# HOMOTOPY VERSUS ISOTOPY: SPHERES WITH DUALS IN 4–MANIFOLDS

ROB SCHNEIDERMAN AND PETER TEICHNER

Abstract. David Gabai recently proved a smooth 4-dimensional "Light Bulb Theorem" in the absence of 2-torsion in the fundamental group. We extend his result to 4–manifolds with arbitrary fundamental group by showing that an invariant of Mike Freedman and Frank Quinn gives the complete obstruction to "homotopy implies isotopy" for embedded 2–spheres which have a common geometric dual. The invariant takes values in an $\mathbb{F}_2$-vector space generated by elements of order 2 in the fundamental group and has applications to unknotting numbers and pseudo-isotopy classes of self-diffeomorphisms. Our methods also give an alternative approach to Gabai's theorem using various maneuvers with Whitney disks and a fundamental isotopy between surgeries along dual circles in an orientable surface.

## 1. Introduction and results

Our starting point is David Gabai's smooth 4-dimensional LBT [3, Thm.1.2]:

**4D-Light Bulb Theorem.** *Let $M$ be a smooth orientable 4–manifold such that $\pi_1 M$ has no elements of order 2. If $R, R' : S^2 \hookrightarrow M$ are smoothly embedded spheres in $M$ which are homotopic, $R \simeq R'$, and have the same geometric dual, then $R$ is smoothly isotopic to $R'$.*

Here a *geometric dual* to a smooth map $R : S^2 \to M$ is a smoothly embedded sphere with trivial normal bundle which intersects $R$ transversely and in a single point. The necessity of the $\pi_1$-condition was shown by Hannah Schwartz in [9] and also follows from Theorem 1.1.

In this paper we extend the above LBT to a version for arbitrary fundamental groups as follows. Fix $f : S^2 \to M$ and a geometric dual $G : S^2 \hookrightarrow M$ for $f$ in a smooth orientable 4–manifold $M$. Consider the following set, measuring "homotopy versus isotopy":

$$\mathcal{R}^G_{[f]} := \{R : S^2 \hookrightarrow M \mid R \simeq f \text{ and } G \text{ is a geometric dual to } R\}/\text{arbitrary isotopies of } R.$$

We shall see that this set is non-empty if and only if the self-intersection invariant of $f$ vanishes, $\mu(f) = 0$, which we will assume from now on. Note that the above set, and the vanishing of $\mu(f)$, only depend on the homotopy class $[f] \in [S^2, M]$.

Let $\mathbb{F}_2 T_M$ be the $\mathbb{F}_2$-vector space with basis $T_M := \{g \in \pi_1 M \mid g^2 = 1 \neq g\}$, the elements of order two (2-torsion). It turns out that the self-intersection invariant for maps $S^3 \looparrowright M^4 \times \mathbb{R}^2$ with transverse double points gives a homomorphism $\mu_3 : \pi_3 M \to \mathbb{F}_2 T_M$ (see Lemma 4.1).

**Theorem 1.1.** *The abelian group $\mathbb{F}_2 T_M$ acts transitively on $\mathcal{R}^G_{[f]}$. If $\mu(f) = 0$ then $\mathcal{R}^G_{[f]} \neq \emptyset$ and each $R : S^2 \hookrightarrow M$ leads to a bijection $\mathcal{R}^G_{[f]} \longleftrightarrow \mathbb{F}_2 T_M/\mu_3(\pi_3 M)$, i.e. the stabilizer of $R \in \mathcal{R}^G_{[f]}$ is always $\mu_3(\pi_3 M) \leq \mathbb{F}_2 T_M$. If $R, R' : S^2 \hookrightarrow M$ represent the same element in $\mathcal{R}^G_{[f]}$ and agree near $G$ then they are isotopic by an isotopy supported away from $G$.*

Gabai's LBT follows: If $\pi_1 M$ contains no 2-torsion then $\mathbb{F}_2 T_M = \{0\}$ and hence $\mathcal{R}^G_{[f]}$ consists of a single isotopy class. In fact, in the second version of his paper Gabai strengthens his result to a “normal form” [3, Thm.1.3] which in the above language translates to saying that there is a surjection $\mathbb{F}_2T_M \twoheadrightarrow \mathcal{R}^G_{[f]}$. Examples where this projection is not injective were given in [9], providing 4-manifolds $M$ for which $\mu_3$ is non-trivial.

**Remark 1.2.** Hannah Schwartz also pointed out the following family of examples, showing that the geometric dual $G$ needs to be common to both spheres: Consider closed, 1-connected manifolds $M_0, M_1$ with indefinite intersection form that are stably diffeomorphic but not diffeomorphic. There are many such examples and we may assume that

$$\varphi : M_0\#S^2 \times S^2 \xrightarrow{\cong} M_1\#S^2 \times S^2$$

is a diffeomorphism which preserves the $S^2\times S^2$-summands homotopically. This last assumption needs the indefiniteness – then the diffeomorphism group acts transitively on hyperbolic summands in $\pi_2$. Consider the spheres $R := S^2 \times p$ and $R' := \varphi(S^2 \times p)$ in $M_1\#S^2 \times S^2$ with geometric duals $p \times S^2$ and $\varphi(p \times S^2)$. Then $R$ and $R'$ are homotopic but can't be isotopic, otherwise the ambient isotopy theorem would lead to a diffeomorphism $M_0 \cong M_1$.

1.A. **Consequences of Theorem 1.1 and its proof.**

**Corollary 1.3.** *There exist 4–manifolds $M$ and $f : S^2 \looparrowright M$ with infinitely many free isotopy classes of embedded spheres homotopic to $f$ (and with common geometric dual). These manifolds also admit infinitely many distinct pseudo-isotopy classes of self-diffeomorphisms.*

These self-diffeomorphisms (carrying one sphere to the other) will be constructed in Lemma 6.1. For example, let $M'$ be any 4–manifold obtained by attaching 2-handles to a boundary connected sum of copies of $S^1 \times D^3$ such that $\mathbb{Z}/2 * \mathbb{Z}/2 \leq \pi_1 M'$. This infinite dihedral group contains infinitely many distinct reflections (which are of order 2). It follows from Theorem 1.1 that there exist infinitely many isotopy classes of spheres homotopic to $p \times S^2$ in $M := M'\#(S^2 \times S^2)$, all with geometric dual $S^2 \times p$.

**Corollary 1.4.** *Concordance implies isotopy for spheres with a common geometric dual.*

**Corollary 1.5.** *If $R, R' : S^2 \hookrightarrow M^4$ have a common geometric dual in $M$ then $R$ and $R'$ are isotopic in $M \times \mathbb{R}$ if and only if they are isotopic in $M$.*

The two results are actually “scholia”, i.e. corollaries to our proof of Theorem 1.1. Namely, we show that the bijections in our theorem are induced by a based concordance invariant $\mathrm{fq}(R, R') \in \mathbb{F}_2T_M/\mu_3(\pi_3M)$ used by Mike Freedman and Frank Quinn in [2, Thm.10.5(2)] and later named fq by Richard Stong [11, p.2].

As reviewed in Section 4, Freedman–Quinn actually use the self-intersection invariant $\mu_3(H) \in \mathbb{F}_2T_M$ of a map $S^2 \times I \looparrowright M \times \mathbb{R} \times I$ with transverse double points obtained by perturbing the track of a based homotopy $H$ between $R$ and $R'$ in $M \times \mathbb{R}$, explaining Corollary 1.5. Stong states that in the quotient $\mathbb{F}_2T_M/\mu_3(\pi_3M)$, the choice of $H$ disappears and gives $\mathrm{fq}(R, R')$. This will be proven in Section 4.B for any two spheres $R, R'$ in $M$ that are based homotopic.

**Corollary 1.6.** *If $R, R' : S^2 \hookrightarrow M^4$ have a common geometric dual and are homotopic then they are isotopic if and only if* $\mathrm{fq}(R, R') = 0$.

This follows from the relation $\mathrm{fq}(t \cdot R, R) = t$ for all $t \in \mathbb{F}_2T_M/\mu_3(\pi_3M)$, between our action and the Freedman–Quinn invariant, see Section 5.A.

For the next scholem we consider a "relative unknotting number" $\mathrm{u}(R, R')$ for homotopic spheres $R, R' : S^2 \hookrightarrow M$: By assumption, there is a sequence of finger moves and Whitney moves that lead from $R$ to $R'$, compare Section 2.A. Let $\mathrm{u}(R, R') \in \mathbb{N}_0$ denote the minimal number of finger moves required in any such homotopy.

In general, this is an extremely difficult invariant to compute, even though we'll see in Lemma 4.6 that one always has the estimate $\mathrm{u}(R, R') \geq |\,\mathrm{fq}(R, R')|$. Here the *support* $|t|$ is the number of non-zero coefficients in $t \in \mathbb{F}_2 T_M$ and for an equivalence class $[t] \in \mathbb{F}_2 T_M / \mu_3(\pi_3 M)$, we let $|[t]|$ be the minimum support of all representatives.

In the presence of a common geometric dual, Michael Klug pointed out that the above estimate becomes an equality, see the last part of Section 6.

**Corollary 1.7.** *For $R, R' \in \mathcal{R}^G_{[f]}$, the relative unknotting number equals the support of the Freedman-Quinn invariant:* $\mathrm{u}(R, R') = |\,\mathrm{fq}(R, R')|$.

Using the 4-manifold $M$ below Corollary 1.3, we see that any (arbitrary large) number is realized as the relative unknotting number between spheres in $M$. This should also be true for knotted spheres in $S^4$ relative to the unknot.

1.B. **An isotopy invariant statement.** Even though the original LBT's in $S^2 \times S^1$ and $S^2 \times S^2$ are extremely well motivated, see [3], categorically oriented readers may find it confusing that our set $\mathcal{R}^G_{[f]}$ is *not isotopy invariant*: If we do a finger move on $R \in \mathcal{R}^G_{[f]}$ that introduces two additional intersection points with the dual $G$, the resulting embedded sphere is isotopic to $R$ but not in $\mathcal{R}^G_{[f]}$ any more. In other words, if one wants isotopy invariant statements, one should not fix a sphere $G$ as in the LBT. We address this problem as follows:

**Definition 1.8.** For fixed $R : S^2 \hookrightarrow M^4$ with fixed geometric dual $G$, consider pairs of embeddings $R', G' : S^2 \hookrightarrow M$ such that:

- $R'$ is *homotopic* to $R$ via $R_s : S^2 \to M$,
- $G'$ is *isotopic* to $G$ via $G_s : S^2 \hookrightarrow M$, and
- $G_s$ is a geometric dual to $R_s$ for each $s \in I$.

Denote by $\mathcal{R}_{R,G}$ the set of isotopy classes of such pairs $(R', G')$, where an *isotopy of a pair* is a pair $(R_s, G_s)$ as above where $R_s$ is in addition an isotopy.

Then an isotopy $R_s$ can be embedded into an ambient isotopy $\varphi_s : M \xrightarrow{\cong} M$ and hence leads to pairs $(R_s, \varphi_s(G))$ that are all equal in $\mathcal{R}_{R,G}$.

**Theorem 1.9.** *The group $\mathbb{F}_2 T_M$ acts transitively on $\mathcal{R}_{R,G}$, with stabilizers $\mu_3(\pi_3 M)$. The Freedman-Quinn invariants* $\mathrm{fq}(R, R') = [\mu_3(R_s)]$ *lead to the inverse of the resulting bijection* $\mathcal{R}_{R,G} \longleftrightarrow \mathbb{F}_2 T_M / \mu_3(\pi_3 M)$.

It turns out that this result is equivalent to Theorem 1.1 above and as a consequence, we won't follow up on it in this paper. For example, to derive Theorem 1.1, we can use Lemma 2.1 to turn any homotopy $R_s$ into one that satisfies the last condition in the above definition (with $G_s = G$ a constant isotopy).

1.C. **Outline of the proof of Theorem 1.1.** Our action of $t = t_1 + \cdots + t_n \in \mathbb{F}_2 T_M$ on $R \in \mathcal{R}^G_{[f]}$ will be defined as follows. First create a generic map $f^t : S^2 \looparrowright M$ by doing $n$ finger moves on $R$ along arcs representing $t_i \in T_M$. There is a collection of $n$ Whitney

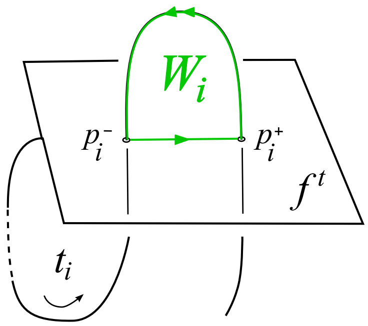


FIGURE 1. The Whitney disks $W$ and $W^t$ pair the same self-intersections $p_i^{\pm}$ of $f^t$. On the boundary $\partial W^t$ differs from $\partial W$ as it departs and approaches from the negative self-intersection $p_i^-$ in different local sheets of $f^t$.

disks $\mathcal{W} \subset M \smallsetminus G$ for $f^t$ which are "inverse" to the finger moves, i.e. the result of doing the Whitney moves along the Whitney disks in $\mathcal{W}$ is isotopic to $R$.

Since $t_i^2 = 1$, we can use $G$ to find a different collection $\mathcal{W}^t \subset M \smallsetminus G$ of $n$ Whitney disks on $f^t$ which induce the same pairings of self-intersections of $f^t$ as $\mathcal{W}$ but which induce different sheet choices for the preimages of the self-intersections, see Figure 1 and Lemma 2.4. The result of doing the Whitney moves on the Whitney disks in $\mathcal{W}^t$ is an embedded sphere denoted by $t \cdot R$, which by construction is homotopic to $R$ and has geometric dual $G$. We'll show in Section 5 that $t \cdot R$ is isotopic to $R$ if and only if $t \in \mu_3(\pi_3 M)$.

The isotopy class of $t \cdot R$ can also be described explicitly without knowing the Whitney collection $\mathcal{W}^t$ by the following *Norman sphere*, built from $f^t$ and $G$ (see Section 3.C). Instead of doing Whitney moves on $\mathcal{W}^t$, the *Norman trick* [5] can be applied to eliminate the self-intersections of $f^t$ by tubing into the dual sphere $G$ along arcs in $f^t$. This operation also involves a choice of local sheets for each self-intersection, and we will show in Section 3.D that $t \cdot R$ is isotopic to the result of applying the Norman trick using the *opposite* local sheets at each negative self-intersection compared to the original finger moves.

Gabai's proof of his LBT in [3] introduces a notion of "shadowing a homotopy by a tubed surface", which uses careful manipulations of several types of tubes and their guiding arcs to control the isotopy class of the result of a homotopy between embeddings. In addition to using the Norman trick, Gabai also works with tubes along framed arcs that extend into the ambient 4–manifold, including the guiding arcs for finger moves.

Our proof of Theorem 1.1, which implies Gabai's LBT, takes a different viewpoint by focussing on the generic sphere $f$ which is the middle level of a homotopy between embeddings $R$ and $R'$, given by finger moves and then Whitney moves. By reversing the finger moves, we see that both these embeddings are obtained from $f$ by sequences of Whitney moves along two collections of Whitney disks. We analyze all choices involved in such collections of Whitney disks and show how they are related to the Freedman–Quinn invariant $\mathrm{fq}(R, R')$.

Our key tool is the relationship between Whitney moves and surgeries on surfaces as shown in Figure 2. Note that the dual curve to $\partial c_W$ on $F$ is a meridional circle to $F$ which bounds a meridional disk. This meridional disk $d$ is usually of not much use since it intersects $F$ (exactly once). However, in the presence of the dual $G$ to $F$, we can tube $d$ into $G$, removing this intersection and obtaining a cap $c_G$ which is disjoint from $F$ and $c_W$, see Figure 8.

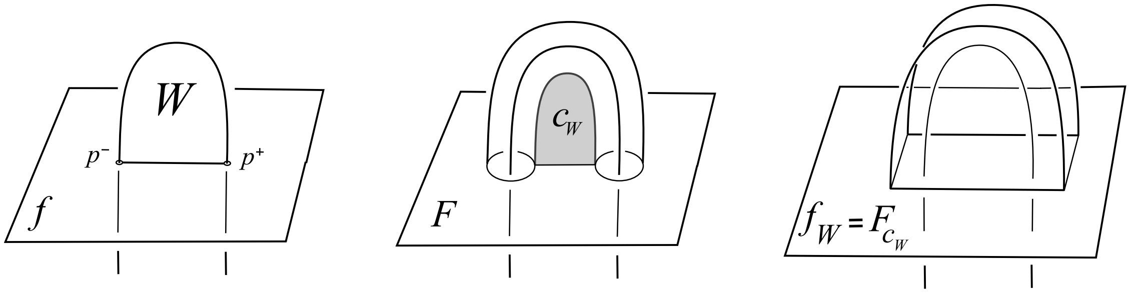


FIGURE 2. Left: $W$ pairing self-intersections $p^{\pm}$ of $f$. Center: The surface $F$ obtained by tubing $f$ to itself admits a cap $c_W$. Right: The result $f_W$ of doing the $W$-move is isotopic to the result $F_{c_W}$ of surgery on $c_W$.

Now we can apply the fundamental isotopy between surgeries along dual curves in an orientable surface showing that surgery on $c_W$ is isotopic to surgery on $c_G$ (see Section 2.G). And if $W'$ is any other Whitney disk for $f$ having the same Whitney circle $\partial W' = \partial W$, then we see that surgery on $c_G$ is also isotopic to surgery on $c_{W'}$. All together, this implies that the Whitney moves on $f$ along $W$, respectively $W'$, give isotopic results $R$, respectively $R'$!

This outline finishes our proof in the very simple case that our collections contain only one Whitney disk and the Whitney circles agree. Multiple Whitney disks in our collections correspond to higher genus capped surfaces and the remaining steps in the argument are "only" about showing independence of Whitney circles. Section 3.D consists of a sequence of lemmas that reduce this dependence only to the choices of sheets at self-intersections whose group elements are of order 2. Fortunately, these are exactly detected by the Freedman–Quinn invariant, finishing our proof.

1.D. **Embedded spheres in other dimensions.** In the forthcoming paper [8], we will give a classification of isotopy versus homotopy for embedded 2–spheres in 5–manifolds. This was inspired by the current results and the fact that the Freedman–Quinn invariant factors through $M^4 \times \mathbb{R}$. For an oriented 5-manifold $N$ and a fixed embedding $F : S^2 \hookrightarrow N^5$, we will describe the set

$$\mathcal{R}_{[F]} := \{R : S^2 \hookrightarrow N^5 \mid R \simeq F\}/\text{isotopy}$$

in purely algebraic terms. If $N^5 = M^4 \times \mathbb{R}$ and $F : S^2 \hookrightarrow M \times 0$ has a geometric dual $G$ in $M$ then our 4d- and 5d-Theorems are related by the following commutative diagram:

$$\begin{tikzcd}
\mathcal{R}^G_{[F]}(M^4) \arrow[r, tail, "i"] \arrow[d, leftrightarrow, "\text{action \&}"', "\text{fq}(-,F)"] & \mathcal{R}_{[F]}(M^4 \times \mathbb{R}) \arrow[d, leftrightarrow, "\text{action \&}"', "\text{fq}(-,F)"] \\
\mathbb{F}_2 T_M/\mu_3(\pi_3 M) \arrow[r, tail] & \mathbb{Z}\pi_1 M/\langle g + g^{-1}, 1, \mu_3(\pi_3 M)\rangle
\end{tikzcd}$$

The inclusion $i$ in the upper row maps onto those spheres $R : S^2 \hookrightarrow M \times 0 \subset M \times \mathbb{R}$ that have $G$ as a geometric dual in $M$. This is quite a "small" subset since the cokernel of the monomorphism on the bottom is the free abelian group whose basis is the quotient of the set $\{g \in \pi_1 M \mid g^2 \neq 1\}$ by the involution $g \mapsto g^{-1}$.

In dimensions $d \neq 4, 5$, homotopy implies isotopy for embeddings $S^2 \hookrightarrow X^d$ in any $d$-manifold $X$. For $d > 5$ this is just general position and otherwise it follows from special features of low dimensional manifolds: For $d = 2$, any embedding of $S^2$ must map onto a component of $X$, so the result is clear.

In dimension $d = 3$ the standard inclusion $S^2 \subset \mathbb{R}^3$ is not isotopic to its pre-composition with a reflection, even though it's (regularly!) homotopic. However, up to this reflection, Laudenbach proved in 1973 [4, Thm.I] that homotopy implies isotopy for spheres in all 3-manifolds. He had to assume the Poincaré conjecture which is known by now.

*Acknowledgements:* It is a pleasure to thank David Gabai and Daniel Kasprowski for helpful discussions. The first author was supported by a Simons Foundation *Collaboration Grant for Mathematicians*, and both authors thank the Max Planck Institute for Mathematics, Bonn, where this work was carried out.

## Contents

1. Introduction and results 1
1.A. Consequences of Theorem 1.1 and its proof 2
1.B. An isotopy invariant statement 3
1.C. Outline of the proof of Theorem 1.1 3
1.D. Embedded spheres in other dimensions 5
2. Preliminaries on surfaces in 4-manifolds 7
2.A. Homotopy classes of surfaces 7
2.B. Self-intersection invariants 9
2.C. Whitney disks and Whitney moves 9
2.D. Sliding Whitney disks 10
2.E. Tubing into the dual sphere 11
2.F. Clean collections of Whitney disks 11
2.G. Capped surfaces and Whitney moves 12
3. New Proof of Gabai's LBT 17
3.A. Choices of sheets, pairings, W-arcs and W-disks 17
3.B. Existence and choices of Whitney disks 18
3.C. Norman spheres 19
3.D. Independence of pairings and Whitney arcs 23
3.E. Double sheet changes 23
3.F. Choice of sheets for double point loops not of order 2 23
3.G. Choice of sheets for trivial double point loops 24
4. The Freedman–Quinn invariant 25
4.A. 3–manifolds in 6–manifolds 25
4.B. The self-intersection invariant for homotopies of 2–spheres in 5–manifolds 27
4.C. Computing the Freedman–Quinn invariant 28
4.D. Singular circles: The origin of the fq invariant 29
5. Proof of Theorem 1.1 31
5.A. The geometric action on $\mathcal{R}^G_{[f]}$ 31
5.B. The stabilizer equals $\mu_3(\pi_3 M)$ 32
5.C. The action is transitive 32
6. Proofs of Corollaries 1.3 and 1.7 32
7. Ambient Morse theory and the $\pi_1$-negligible embedding Theorem 33
References 39

## 2. Preliminaries on surfaces in 4-manifolds

Statements will hold in smooth and locally flat topological categories, and smoothing of corners will be assumed without mention during cut-and-paste operations on surfaces. Orientations will usually be assumed and suppressed, as will choices of basepoints and whiskers.

In the smooth category, a *generic* map, written $f : \Sigma^2 \looparrowright M^4$, is a smooth map which is an embedding, except for a finite number of transverse double points. This means that there are coordinates on $\Sigma$ and $M$ such that $f$ looks locally like the inclusion $\mathbb{R}^2 \times \{0\} \subset \mathbb{R}^4$ or like a transverse double point $\mathbb{R}^2 \times \{0\} \cup \{0\} \times \mathbb{R}^2 \subset \mathbb{R}^4$.

If $\Sigma$ is closed, generic maps, sometimes also referred to as generic immersions, are open and dense in the Whitney topology on $C^\infty(\Sigma, M)$. In particular, any smooth map is homotopic to a generic map. By topological transversality [2, Chap.9], every continuous map of a surface to a topological 4–manifold is homotopic to one with these two kinds of local behavior, that's one reason why everything we'll do smoothly continues to hold in the topological category. Alternatively, we could use smoothing theorems (away from points in $M$) to put ourself into a smooth setting if needed.

2.A. **Homotopy classes of surfaces.** We will use the following fact about homotopy classes $[\Sigma, M]$ of maps $f : \Sigma \to M$ when $\Sigma$ and $M$ are oriented: The inclusion of generic maps into all smooth maps induces a bijection

$$\{f : \Sigma \looparrowright M \mid \#(f \pitchfork f) = 0\}/\{\text{isotopies, finger moves, Whitney moves}\} \longleftrightarrow [\Sigma, M]$$

where $\#(f \pitchfork f) \in \mathbb{Z}$ denotes the signed sum of double points of $f$. Note that $\#(f \pitchfork f)$ can be changed arbitrarily by (non-regular) cusp homotopies and in the following, we'll always tacitly assume that this has been done such that $\#(f \pitchfork f) = 0$.

In the smooth category, the bijection follows from Whitney's classification of singularites [14] of maps from 3-manifolds to 5-manifolds. It implies that the singularities of a generic track of a homotopy $H$ consist only of finger moves, Whitney moves and cusps. These arise at finitely many times $t \in I$, when $H_t : \Sigma \to M$ is *not* generic but has a tangency (increasing respectively decreasing the double point set by a pair with opposite signs) or when the rank of its derivative drops (creating a cusp where one double point appears or disappears). If $\#(f \pitchfork f) = \#(f' \pitchfork f')$ then the cusps arising in $H$ can be cancelled in pairs, see [2, Chap.1] for the topological case.

In the setting of the LBT, finger moves in a generic homotopy from $R$ to $R'$ having common geometric dual $G$ may be assumed to be disjoint from $G$ since finger moves are supported near their guiding arcs. By the following lemma, the Whitney moves in such a homotopy may also be assumed to be disjoint from $G$ because one easily finds a preliminary isotopy that makes $R$ and $R'$ agree near $G$. This is also [3, Lem.6.1] where the 3D-LBT is used in the proof. For the convenience of the reader, and for completeness, we give an elementary argument.

**Lemma 2.1.** *If $R, R' : S^2 \hookrightarrow M$ agree near a common geometric dual $G$ and are homotopic in $M$ then there exists a finite sequence of isotopies, finger moves and Whitney moves in $M \smallsetminus G$ leading from $R$ to $R'$.*

*Proof.* We first show that $R, R'$ are base point preserving homotopic, noting that they both send a base-point $z_0 \in S^2$ to $z = R \cap G = R' \cap G$ and hence represent elements $[R], [R'] \in$

$\pi_2(M,z)$. Any free homotopy $H$ from $R$ to $R'$ identifies $[R']$ with $g\cdot[R]$, where the loop $H(z_0\times I)$ represents $g\in\pi_1(M,z)$ and we use the $\pi_1$-action on $\pi_2$.

Now take a free homotopy $H$ that is transverse to $G\subset M$ and consider the submanifold $L:=H^{-1}(G)\subset S^2\times I$. $L$ is a 1-manifold with boundary $z_0\times\{0,1\}$ since $R$ and $R'$ intersect $G$ exactly in $z\in M$. This implies that $L$ has a component $L_0$ which is homotopic (in $S^2\times I$) to $z_0\times I$ rel endpoints. As a consequence, the above group element $g$ is also represented by $H(L_0)\subset G\cong S^2$ and hence $[R]=[R']$.

We could next work on the other components of $L$ but prefer the following homological argument. Removing an open normal bundle of $G$ leads to a 4-manifold $W:=M\smallsetminus\nu G$ with a new boundary component $\partial_0 W\cong S^2\times S^1$. $W$ contains two embedded disks $D$ and $D'$ with the same boundary in $\partial_0 W$. These disks complete to the spheres $R$ and $R'$ when adding $G$ back into the 4-manifold.

We claim that $D$ and $D'$ are homotopic rel boundary in $W$ by the homological argument below. Granted this fact, we see from the above discussion that there is also a regular homotopy rel boundary from $D$ to $D'$ in $W$. Approximating it by a generic map we obtain the desired type of homotopy in $M\smallsetminus G$.

To show that $D$ and $D'$ are homotopic rel boundary in $W$, it suffices to show that the glued up sphere $S:=D\cup_\partial D'$ is null homotopic in $W$. Since $R$ intersects $G$ in a single point, it follows from Seifert-van Kampen that the inclusion induces an isomorphism $\pi_1 W\cong\pi_1 M$, with base-points taken on $\partial_0 W$. The long exact sequence of the pair $(M,W)$ for homology with coefficients in $\mathbb{Z}\pi_1 W$ gives exactness for

$$H_3(M,W;\mathbb{Z}\pi_1 W)\longrightarrow H_2(W;\mathbb{Z}\pi_1 W)\longrightarrow H_2(M;\mathbb{Z}\pi_1 M).$$

The Hurewicz isomorphism identifies the map on the right hand side with $\pi_2 W\longrightarrow\pi_2 M$ which sends $S$ to zero by our conclusion on $R,R'$ being based homotopic. By excision and Lefschetz duality,

$$H_3(M,W;\mathbb{Z}\pi_1 W)\cong H_3(S^2\times D^2,S^2\times S^1;\mathbb{Z}\pi_1 W)\cong H^1(S^2\times D^2;\mathbb{Z}\pi_1 W)=0$$

which implies that $[S]=0$. $\square$

We note that Lemma 2.1 is the reason why free (versus based) homotopy and isotopy agree in the presence of a common dual. In particular, why we don't have to divide out by the conjugation action of $\pi_1 M$ in Theorem 1.1. In our forthcoming paper on 2-spheres in 5-manifolds [8] this difference will be relevant and fully exploited.

In the rest of the paper, we will turn a sequence of finger moves and Whitney moves as in Lemma 2.1 into an isotopy, provided the Freedman–Quinn invariant vanishes. If $f:S^2\looparrowright M$ is the *middle level* of such a sequence, i.e. the result of all finger moves on $R$, then there are two *clean collections of Whitney disks* for $f$ in $M$: One collection $\mathcal{W}$ reverses all the finger moves and leads back to $R=f_{\mathcal{W}}$, and the other collection $\mathcal{W}'$ does the interesting Whitney moves to arive at $R'=f_{\mathcal{W}'}$.

Thus the triple $(f,\mathcal{W},\mathcal{W}')$ represents the entire homotopy from $R$ to $R'$ up to isotopy. By construction, the two collections of Whitney disks are *clean* in the sense of Definition 2.3 which formalizes the above discussion. In particular, since the result of Lemma 2.1 is a homotopy in the complement of $G$, the notion of clean Whitney disk will include disjointness from $G$. Then all our maneuvers will stay in the complement of $G$, explaining the last sentence in Theorem 1.1.

2.B. **Self-intersection invariants.** Let $M$ be a smooth oriented 4–manifold and let $f : S^2 \looparrowright M$ be a generic sphere with a whisker from the base point of $M$ to $f$. A loop in $f(S^2)$ that changes sheets exactly at one self-intersection $p$ is called a *double point loop* at $p$. After choosing an orientation of the double point loop, it determines an element $g \in \pi_1 M$ associated to $p$. The orientation of a double point loop corresponds to a *choice of sheets* at $p$, i.e. a choice of a point $x \in f^{-1}(p)$ that is the starting point of the preimage of the loop.

The self-intersection invariant $\mu(f) \in \mathbb{Z}[\pi_1 M]/\langle g - g^{-1}\rangle$ is defined by summing the group elements represented by double point loops of $f$, with the coefficients coming from the usual signs determined by the orientation of $M$. The relations $g - g^{-1} = 0$ in the integral group ring account for the above choices of sheets.

Then $\mu(f)$ is invariant under generic regular homotopies of $f$ and changes by $\pm 1$ under a cusp homotopy. Therefore, taking $\mu(f)$ in a further quotient that also sets the identity element $1 \in \pi_1 M$ equal to 0 makes $\mu(f)$ invariant under arbitrary homotopies of $f$. In the literature the adjective "reduced" is sometimes attached to the self-intersection invariant $\mu$ when including this relation $1 = 0$ in the target.

The analogous reduced self-intersection invariant defined for generic 3-spheres in 6–manifolds will be relevant in Section 4.

2.C. **Whitney disks and Whitney moves.** Suppose that a pair $p^{\pm}$ of oppositely-signed self-intersection points of $f : S^2 \looparrowright M$ have equal group elements for some choices of sheets at $p^+$ and $p^-$. Then the pair $p^{\pm}$ admits an embedded null-homotopic *Whitney circle* $\alpha \cup \beta = f(a) \cup f(b)$ for disjointly embedded arcs $a$ and $b$ joining the preimages $x^+, y^+$ and $x^-, y^-$ of $p^+$ and $p^-$, as in Figure 3. Such $\alpha$ and $\beta$ are called *Whitney arcs*.

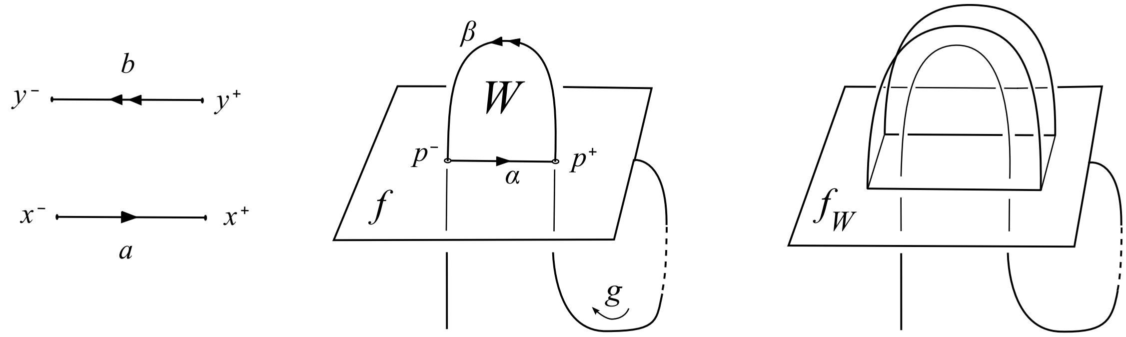


FIGURE 3. Left: In the domain of $f$. Center: The horizontal sheet of $f$ appears in the 'present' as does the Whitney disk $W$, and the other sheet of $f$ appears as an arc which is understood to extend into 'past and future', with the dashed part indicating where $f$ extends outside the pictured 4–ball neighborhood of $W$ in $M$. Right: After the Whitney move guided by $W$.

The center of Figure 3 also shows a *Whitney disk* $W$ with boundary $\partial W = \alpha \cup \beta$ pairing self-intersections $p^{\pm}$ with group element $g \in \pi_1 M$. The right side of Figure 3 shows the result $f_W$ of doing a *Whitney move* on $f$ guided by $W$, which is an isotopy of one sheet of $f$, supported in a regular neighborhood of $W$, that eliminates the pair $p^{\pm}$. Combinatorially, $f_W$ is constructed from $f$ by replacing a regular neighborhood of one arc of $\partial W$ with a *Whitney bubble* over that arc. This Whitney bubble is formed from two parallel copies of $W$ connected by a curved strip which is normal to a neighborhood of the other arc. Figure 3 shows $f_W$

using a Whitney bubble over $\alpha$. Although both these descriptions of $f_W$ involve a choice of arc of $\partial W$, up to isotopy $f_W$ is independent of this choice.

The construction of an embedded Whitney bubble requires that $W$ is *framed* (so that the two parallel copies used above do not intersect each other), and Whitney disks which do not satisfy the framing condition are called *twisted* (see eg. [7, Sec.7A]).

2.D. **Sliding Whitney disks.** We describe here an operation that "slides" Whitney disks over each other. This maneuver changes the Whitney arcs while preserving the isotopy class of the results of the Whitney moves, and will be used in the proof of the key Proposition 2.11.

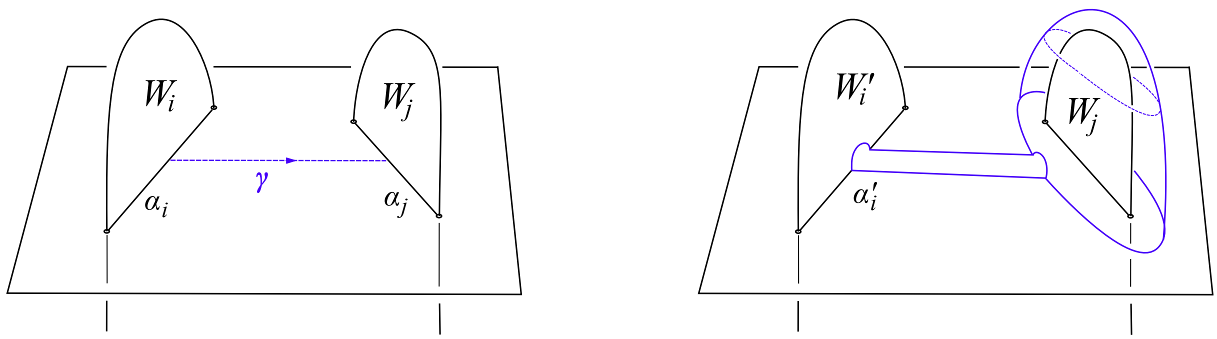


FIGURE 4. Left: A path $\gamma$ guiding a slide of $W_i$ over $W_j$. Right: The result $W_i'$ of sliding $W_i$ contains the (blue) Whitney bubble $B_{\alpha_j}$ over $\alpha_j$.

Let $W_i$ and $W_j$ be two Whitney disks on $f$, and let $\gamma$ be an embedded path in $f$ from $\alpha_i \subset \partial W_i$ to $\alpha_j \subset \partial W_j$ such that the interior of $\gamma$ is disjoint from any self-intersection of $f$ or Whitney arcs on $f$. Denote by $W_i'$ the result of boundary-band-summing $W_i$ into a Whitney bubble $B_{\alpha_j}$ over $\alpha_j$ by a half-tube along $\gamma$ as in Figure 4. We say that $W_i'$ is the result of *sliding* $W_i$ *over* $W_j$.

To see that $f_{\{W_i',W_j\}}$ is isotopic to $f_{\{W_i,W_j\}}$, just observe that $W_i'$ becomes isotopic to $W_i$ after doing the $W_j$-Whitney move. To see this in the coordinates of Figure 4, note that doing the $W_j$-Whitney move would either replace a horizontal disk of $f$ inside $B_{\alpha_j} \subset W_i'$ by a smaller Whitney bubble over $\alpha_j$, or would leave the same horizontal disk free of intersections by adding a Whitney bubble over $\beta_j$ to the other sheet of $f$. So $W_i'$ isotopes back to $W_i$ across the smaller bubble or the horizontal disk.

Either of $\alpha_i$ and $\beta_i$ can be slid over either of $\alpha_j$ or $\beta_j$, and the isotopy class of the results of Whitney moves will be preserved as long as $i \neq j$. This sliding operation can be iterated:

**Lemma 2.2.** *If a collection $\mathcal{W}'$ of Whitney disks on $f$ is the result of performing finitely many slides ($i \neq j$) on a collection $\mathcal{W}$, then $f_{\mathcal{W}'}$ is isotopic to $f_{\mathcal{W}}$.* □

Regarding the $i = j$ case, one can indeed slide $W_i$ over itself using a band from $\alpha_i \subset \partial W_i$ to the boundary of a Whitney bubble $B_{\beta_i}$ over $\beta_i$, and the result will still be a clean Whitney disk. We don't believe that such a *self-slide* will preserve the isotopy class of $f_{\mathcal{W}}$ in general (as it does in Lemma 2.2). However, it will follow from Lemma 2.10 that this self-sliding does indeed preserve the isotopy class of the result of the Whitney move in our current setting where $f$ is a sphere with a geometric dual.

2.E. **Tubing into the dual sphere.** For $G$ a geometric dual to $f$, a transverse intersection point $r$ between $f$ and a surface $D$ can be eliminated by tubing $D$ into $G$. This is known as the *Norman trick* [5] and is the main reason why dual spheres are so useful. Here "tubing $D$ into $G$" means taking an ambient connected sum of $D$ with a parallel copy $G'$ of $G$ via a tube (an annulus) of normal circles over an embedded arc in $f$ that joins $r$ with an intersection point between $f$ and $G'$, see Figure 5. Note that in the case that $D = f$ this operation involves a choice of which local sheet of $r$ to connect into.

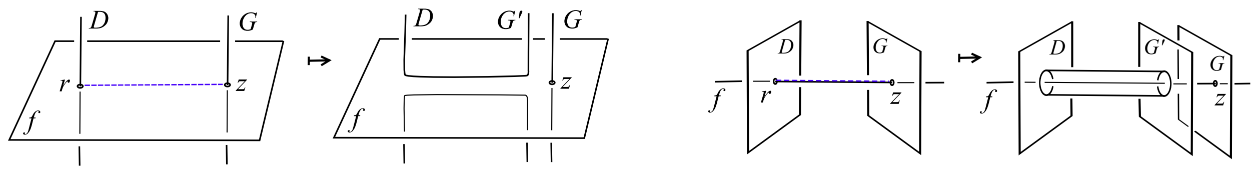


FIGURE 5. Two views of the 'tubing into $G$' operation to eliminate $r \in f \pitchfork D$, guided by a (blue dashed) path from $r$ to $z = f \cap G$.

There are infinitely many pairwise disjoint copies of $G$ intersecting a small neighborhood around $z = f \cap G$ in $f$, so this procedure can be applied to eliminate any number of such intersections without creating new ones as long as appropriate guiding arcs for the tubes can be found. By varying the radii of the tubes, the guiding arcs can be allowed to intersect while keeping the tubes disjointly embedded.

2.F. **Clean collections of Whitney disks.** Recall that for $f : S^2 \looparrowright M^4$, the vanishing of the self-intersection invariant

$$\mu(f) = 0 \in \mathbb{Z}[\pi_1 M]/\langle g - g^{-1} \rangle$$

is equivalent to the existence of choices of sheets so that all double points of $f$ can be arranged in pairs admitting null-homotopic Whitney circles (this statement is independent of the chosen whisker for $f$).

**Definition 2.3.** A *clean collection of Whitney disks* for $f : S^2 \looparrowright M$ is a collection of Whitney disks that pair all double points of $f$ and are framed, disjointly embedded, with interiors disjoint from $f$. In the presence of a dual sphere $G$ for $f$, this notion of a clean collection also includes the disjointness of the Whitney disks from $G$.

Each Whitney disk in a clean collection is called a *clean Whitney disk.*

**Lemma 2.4.** *If $f : S^2 \looparrowright M$ admits a geometric dual $G$, any collection of disjointly embedded Whitney circles that are null-homotopic in $M$ extends to a clean collection of Whitney disks.*

*Proof.* Start with a collection of generic disks $W_i$ bounded by the given null-homotopic Whitney circles that may intersect $G$, may be twisted, and may have interior intersections with $f$ and each other.

Note that the complement in $S^2$ of the union of the preimages $\{a_i, b_i\}$ of the Whitney circles is connected, and that there exist disjointly embedded tube-guiding paths in the complement of the Whitney circles between any number of isolated points and points near $z$.

We describe how to modify the $W_i$ relative their boundaries, without renaming them as changes are made:

First of all, each $W_i$ can be made disjoint from $G$ by tubing $W_i$ into parallel copies of $f$ along disjoint arcs in $G$. Since $f$ is immersed with possibly non-trivial normal bundle, this tubing operation is in general more traumatic than the "tubing into $G$" operation described in section 2.E and creates interior intersections between the $W_i$ and $f$, as well as intersections among the $W_i$.

Next, the intersections and self-intersections among the $W_i$ can be eliminated by pushing each such point down into $f$ by a finger move, and boundary-twists make the $W_i$ framed [2, Chap.1.3], both at the cost of only creating more interior intersections between Whitney disks and $f$.

Finally, the interiors of the $W_i$ can be made disjoint from $f$ by tubing the $W_i$ into $G$ along disjoint paths in $f$. Since $G$ is embedded and has trivial normal bundle the $W_i$ are still framed and disjoint from $G$, i.e. they form a clean collection of Whitney disks $\mathcal{W}$ bounded by the original Whitney circles. □

**Remark 2.5.** The proof of Lemma 2.4 shows that if any subcollection of Whitney circles bound clean Whitney disks, then these same Whitney disks can be extended to a clean collection of Whitney disks by applying the construction to the remaining Whitney circles.

For any given collection $\mathcal{W}$ of clean Whitney disks we denote by $D_z \subset f(S^2)$ a small embedded disk around $z = f(S^2) \cap G$ such that each point in $D_z$ intersects a parallel of $G$ disjoint from $\mathcal{W}$. The radius of $D_z$ is less than the minimum of the radii of the finitely many normal tubes around arcs in $G$ used in the first step of the proof of Lemma 2.4, but our modifications of Whitney disk collections will only use the existence of $D_z$ not its diameter.

The minimum of the radii of the finitely many normal tubes around arcs in $f$ used in the last step of the proof of Lemma 2.4 gives a uniform lower bound on the distance between $f$ and the complements of small boundary collars of all Whitney disks in $\mathcal{W}$. Subsequent modifications of $\mathcal{W}$ by tubing into $G$ along $f$ will always be assumed to use tubes of radius less than this bound, so as long as tubes are away from Whitney disk boundaries the tubes' interiors will be disjoint from Whitney disk interiors.

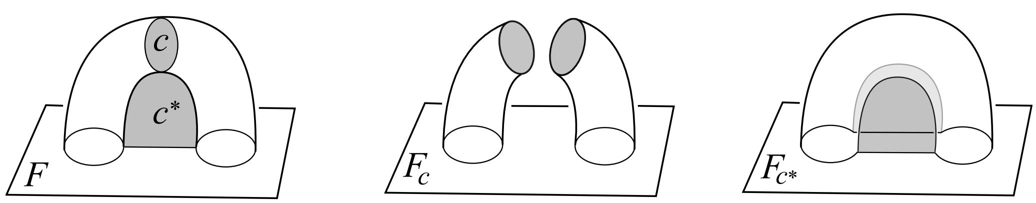


FIGURE 6. A small neighborhood in $\mathbb{R}^3$ of $F \cup c \cup c^*$ on the left is diffeomorphic to $D^2 \times I$, in a way that $D^2 \times \{0\} \cong F_c$ and $D^2 \times \{1\} \cong F_{c^*}$. Hence the two surgeries in the center and right are isotopic

2.G. **Capped surfaces and Whitney moves.** A *cap* on a generic orientable surface $F$ in $M$ is a 0-framed embedded disk $c$ such that the boundary $\partial c$ is a non-separating simple closed curve in $F$, and the interior of $c$ is disjoint from $F$. Two caps on $F$ are *dual* if their boundaries intersect in a single point and their interiors are disjoint. For a collection $\mathcal{C}$ of disjoint caps on $F$, denote by $F_{\mathcal{C}}$ the result of surgering $F$ using all the caps of $\mathcal{C}$. Two such collections $\mathcal{C}$ and $\mathcal{C}^*$ are called *dual* if their boundary curves intersect geometrically $\delta_{ij}$.

The following lemma can be proved by considering an isotopy of a standard model in 3-space that passes through the symmetric surgery on both sets of caps (see Figure 6 and [2, Sec.2.3]):

**Lemma 2.6.** *If $\mathcal{C}$ and $\mathcal{C}^*$ are dual collections of caps on $F$ then $F_{\mathcal{C}}$ is isotopic to $F_{\mathcal{C}^*}$.*

Lemma 2.6, together with the presence of the geometric dual $G$, yields the following simple but useful correspondence between Whitney moves and surgeries:

Let $W$ be a clean Whitney disk on $f$ with $\partial W = \alpha \cup \beta$ (possibly one of a collection $\mathcal{W}$ of Whitney disks on $f$), and let $F : T^2 \looparrowright M$ be the result of tubing $f$ to itself along $\beta$. Observe that a cap $c_W$ on $F$ can be constructed from $W$ by deleting a small boundary collar near $\beta$, and $F_{c_W}$ is isotopic to $f_W$ (Figure 7).

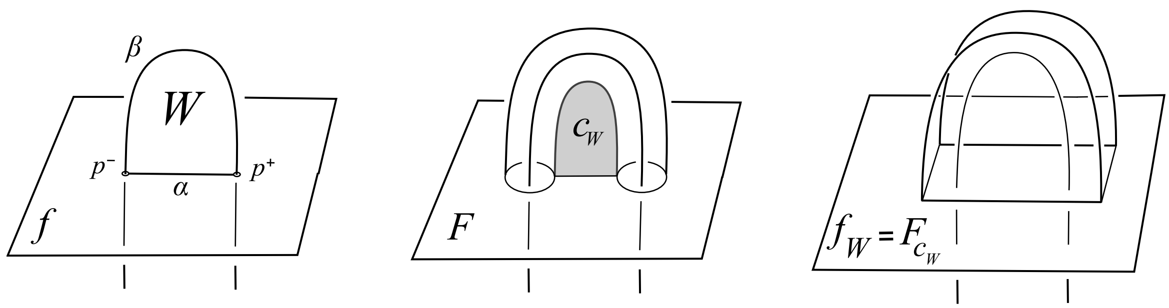


FIGURE 7

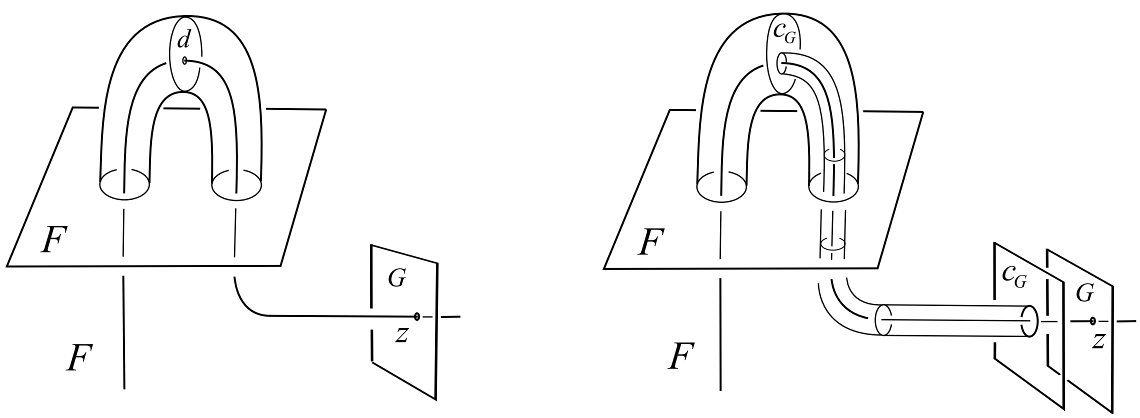


FIGURE 8

Now we construct a cap $c_G$ on $F$ which is dual to $c_W$. Start with a meridional disk $d$ to $F$ which has a single transverse intersection $r = d \pitchfork F \in \beta$ and $\partial d \subset F$ (Figure 8 left). Note that $G$ is a geometric dual to $F$. Then $c_G$ is the result of eliminating $r$ by tubing $d$ into $G$ along an embedded arc in $F$, disjoint from $c_W$ and $\partial d$ (and any other Whitney disks), running from $r$ to a point where a parallel copy of $G$ intersects $F$, see right Figure 8. Such an embedded arc exists since the complement of $\partial W$ is connected (as is the complement of $\partial \mathcal{W}$). Since $c_G$ and $c_W$ are dual caps, Lemma 2.6 gives:

**Lemma 2.7.** *If $F$ is the result of tubing $f$ to itself along one Whitney arc of a clean Whitney disk $W$, and $c_G$ is a cap on $F$ gotten by tubing a meridional disk dual to the Whitney arc into $G$ as above, then $f_W$ is isotopic to $F_{c_G}$.* $\square$

So if two Whitney disks $W$ and $W'$ on $f$ have equal Whitney circles $\partial W = \partial W'$, then $f_W$ is isotopic to $f_{W'}$ since each is isotopic to surgery $F_{c_G}$ on a common dual cap $c_G$ to both of the caps $c_W$ and $c_{W'}$ as in Lemma 2.7. And since the complement in $f$ of the Whitney circles of a clean collection of Whitney disks is connected we have:

**Lemma 2.8.** *If $\mathcal{W}$ and $\mathcal{W}'$ are clean collections of Whitney disks for the self-intersections of $f$ such that $\partial\mathcal{W} = \partial\mathcal{W}'$, then $f_{\mathcal{W}}$ is isotopic to $f_{\mathcal{W}'}$.* □

**Lemma 2.9.** *For the Whitney circles $\mathcal{A} = \partial\mathcal{W}$ of a clean collection $\mathcal{W} = \cup_i W_i$ of Whitney disks as in Definition 2.3, consider $\mathcal{A}'$ which is the result of band summing a Whitney arc $\alpha_i \subset \partial W_i$ into a parallel of $\partial D_z$ along an arc $\gamma$ with interior disjoint from $\mathcal{A}$ as in the left-most and right-most pictures in Figure 9. Then there exists a clean collection of Whitney disks $\mathcal{W}'$ with $\partial\mathcal{W}' = \mathcal{A}'$ and such that $f_{\mathcal{W}}$ is isotopic to $f_{\mathcal{W}'}$.*

*Proof.* We break up the band sum operation into the three steps illustrated in Figure 9: Guided by $\gamma$, modify $\alpha_i$ by pushing a subarc slightly across $\partial D_z$, and extend this isotopy to a collar of $W_i$. The isotopy class of $f_{\mathcal{W}}$ is unchanged since the collection $\mathcal{W}$ changes by isotopy due to the disjointness of $\gamma$ from $\mathcal{A}$.

Now delete from $\alpha_i$ the small (dashed) arc which is the intersection of $\alpha_i$ with the interior of $D_z$, and eliminate the oppositely signed self-intersections of $f$ that were paired by $W_i$ by tubing $f$ along the resulting pair of arcs into two oppositely oriented copies of $G$ which intersect $\partial D_z$ at the arcs' endpoints. See the second picture from the left in Figure 9.

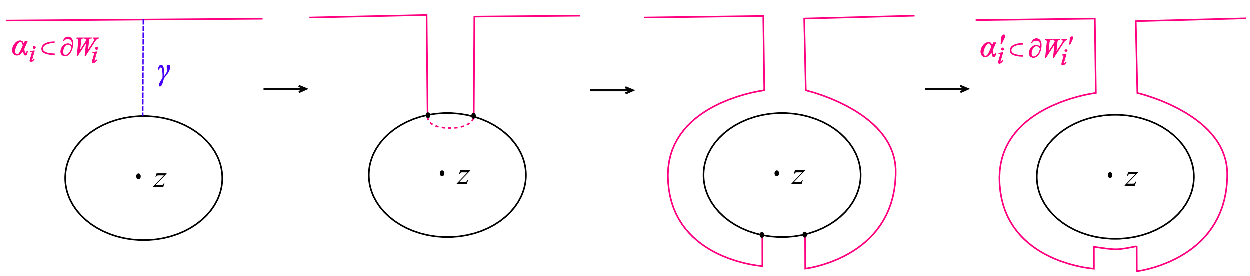


FIGURE 9

This yields an immersed sphere $f^G_\gamma$ which admits the clean collection of Whitney disks $\mathcal{V} := \mathcal{W} \smallsetminus W_i$. Note that by construction $f^G_\gamma$ is also the result of tubing $f$ to itself along the $\alpha_i$ that had been pushed into $D_z$ and then surgering the tube along a cap formed from a parallel copy of $G$ near where $\gamma$ meets $\partial D_z$. It follows from Lemma 2.7 that $f_{W_i}$ is isotopic to $f^G_\gamma$. Hence, $f_{\mathcal{W}}$ is isotopic to $(f^G_\gamma)_{\mathcal{V}}$.

Next, change $f^G_\gamma$ by an isotopy which moves the two tubes and the two parallels of $G$ contained in $f^G_\gamma$ in opposite directions around $\partial D_z$ as shown in the third picture from the left in Figure 9. After this isotopy $f^G_\gamma$ still admits $\mathcal{V}$, and the isotopy class of $(f^G_\gamma)_{\mathcal{V}}$ is unchanged.

Now (re)connect the endpoints of the two guiding arcs of the tubes near the short subarc of $\partial D_z$ between the endpoints to get a single arc $\alpha_i' := \alpha_i +_\gamma \partial D_z$ which is isotopic to the result of taking the band sum of $\alpha_i$ with $\partial D_z$ along $\gamma$ (see the right-most picture in Figure 9). The resulting embedded Whitney circle $\alpha_i' \cup \beta_i$ is null-homotopic and disjoint from $\partial\mathcal{V}$, so by Lemma 2.4 there exists a collection $\mathcal{W}'$ of Whitney disks with $\partial\mathcal{W}' = \alpha_i' \cup \beta_i \cup \partial\mathcal{V}$. As per Remark 2.5, the proof of Lemma 2.4 fixes $\mathcal{V}$ while constructing a clean Whitney disk $W_i'$ bounded by $\alpha_i' \cup \beta_i$ in the complement of $\mathcal{V}$, so we have $\mathcal{W}' = W_i' \cup \mathcal{V}$.

It follows again by Lemma 2.7 that $f_\gamma^G$ is isotopic to $f_{W_i'}$, since $f_\gamma^G$ is isotopic to the result of tubing $f$ to itself along $\alpha_i'$ and then surgering a cap formed from a copy of $G$ near where the guiding arcs were reconnected. Hence $f_{\mathcal{W}'}$ is isotopic to $(f_\gamma^G)_{\mathcal{V}}$, and we see that $f_{\mathcal{W}}$ and $f_{\mathcal{W}'}$ are isotopic. □

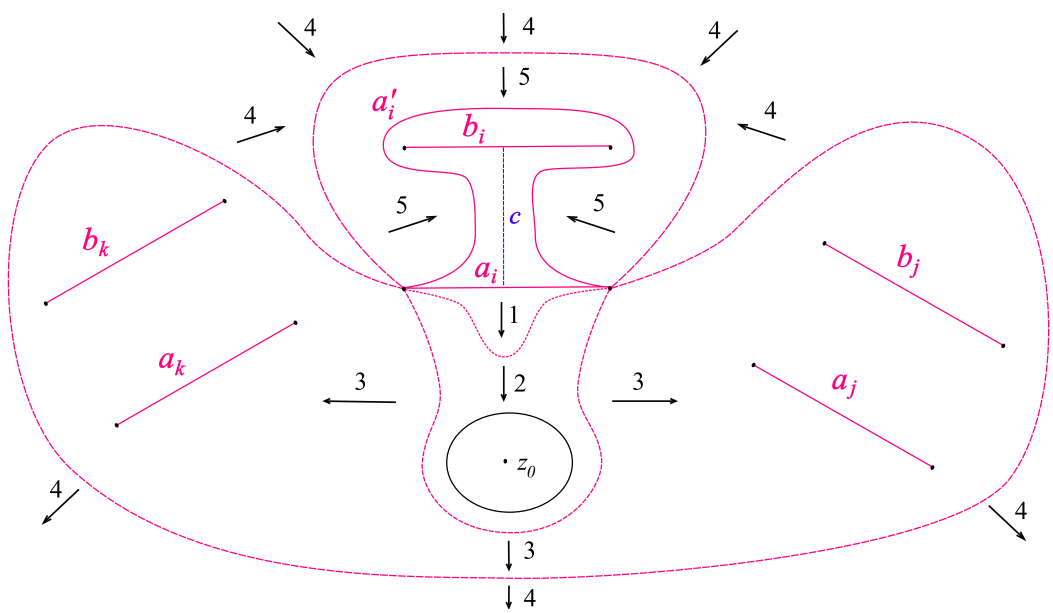


FIGURE 10

**Lemma 2.10.** *For a clean Whitney disk collection $\mathcal{W}$ on $f : S^2 \looparrowright M^4$ with geometric dual $G$, if $\mathcal{W}'$ is gotten from $\mathcal{W}$ by sliding a Whitney disk over itself then $f_{\mathcal{W}'}$ is isotopic to $f_{\mathcal{W}}$.*

*Proof.* Let $\alpha_i$ be the Whitney arc of $\partial W_i = \alpha_i \cup \beta_i$ that is slid over $\beta_i$ to become $\alpha_i' \subset \partial W_i' = \alpha_i' \cup \beta_i$. Referring to Figure 10, consider the following five steps describing in the domain an isotopy of $\alpha_i = f(a_i)$ to $\alpha_i' = f(a_i')$:

Step 1 and Step 2 isotope $\alpha_i$ towards and then across $D_z = f(D_{z_0})$, as in Lemma 2.9. After these first two steps of the isotopy the union of the resulting new arc $\alpha_i^2$ with the original $\beta_i$ admits a clean Whitney disk $W_i^2$, and replacing $W_i$ by $W_i^2$ in $\mathcal{W}$ yields a clean collection $\mathcal{W}^2$ such that $f_{\mathcal{W}^2}$ is isotopic to $f_{\mathcal{W}}$ by Lemma 2.9.

Step 3 then uses the Whitney disk sliding operation of section 2.D to push $\alpha_i^2$ across all the $\alpha_j$ and $\beta_j$ Whitney arcs of the Whitney disks $W_j$ for $j \neq i$ by sliding $W_i^2$ twice over each of these Whitney disks (once each for $\alpha_j$ and $\beta_j$). Taking the resulting Whitney disk $W_i^3$ as a replacement for $W_i^2$ in $\mathcal{W}^2$ yields $\mathcal{W}^3$, with $f_{\mathcal{W}^3}$ isotopic to $f_{\mathcal{W}^2}$ by Lemma 2.2.

Finally, Steps 4 and 5 isotope a collar of $W_i^3$ around the 2-sphere until the Whitney disk boundary arc ends up as the band sum $\alpha_i'$ of the original $\alpha_i$ with the boundary of a Whitney bubble over $\beta_i$. This 5-step construction yields $\mathcal{W}^5$ with $W_i^5 \in \mathcal{W}^5$ having boundary $\alpha_i' \cup \beta_i$ and $f_{\mathcal{W}}$ isotopic to $f_{\mathcal{W}^5}$. Now form $\mathcal{W}'$ from $\mathcal{W}^5$ by replacing the Whitney disk $W_i^5$ resulting from this construction with the Whitney disk $W_i'$ gotten by sliding $\alpha_i$ across $\beta_i$ which has the same boundary. By Lemma 2.8 we get that $f_{\mathcal{W}}$ is isotopic to $f_{\mathcal{W}'}$. □

We come to our most useful geometric result for $f : S^2 \looparrowright M$ with geometric dual $G$:

**Proposition 2.11.** *If $\mathcal{W}$ and $\mathcal{W}'$ are clean collections of Whitney disks on $f$ such that for each $i$, $W_i \in \mathcal{W}$ and $W'_i \in \mathcal{W}'$ share a common Whitney arc $\beta_i = \beta'_i$, then $f_{\mathcal{W}}$ is isotopic to $f_{\mathcal{W}'}$.*

*Proof.* We first prove the simplest case of the statement: If $W$ and $W'$ are Whitney disks on $f$ which share a common Whitney arc $\beta = \beta'$, then $f_W$ is isotopic to $f_{W'}$.

The proof will proceed as in the setting of Lemma 2.7, but because here we have two Whitney disks with possibly $\alpha \neq \alpha'$ we may need to apply the sliding maneuver of Section 2.D to create a tube-guiding arc to $z$ for cleaning up the meridional cap.

Let $F$ be the surface resulting from tubing $f$ to itself along the common Whitney arc $\beta = \beta'$ of $\partial W$ and $\partial W'$. Deleting small boundary collars of $W$ and $W'$ near $\beta$ yields caps $c_W$ and $c_{W'}$ for $F$ as in Figure 7, but with $\partial c_{W'}$ wandering off into the "horizontal" part of $F$ corresponding to $\alpha' \neq \alpha$. By Lemma 2.7, $F_{c_W}$ is isotopic to $f_W$, and $F_{c_{W'}}$ is isotopic to $f_{W'}$.

As in the setting of Lemma 2.7, we want to construct a cap $c_G$ for $F$ such that $c_G$ is dual to both $c_W$ and $c_{W'}$. Then by Lemma 2.7 it will follow that each of $f_W$ and $f_{W'}$ is isotopic to $F_{c_G}$.

The construction of $c_G$ starts as in Figure 8: We want to clean up a meridional disk $d$ to $F$ which has a single transverse intersection $r = d \pitchfork F \in \beta$ and $\partial d \subset F$ by tubing $d$ into $G$. But now we have to find an embedded path from $r$ to $z = G \cap F$ that is disjoint from *both* $\partial c_W$ and $\partial c_{W'}$.

If $r$ and $z$ lie in the same connected component of $F \smallsetminus (\partial c_W \cup \partial c_{W'})$ then there is no problem. We can eliminate $r$ by tubing $d$ into $G$ along an embedded path in $F$ running from $r$ to a point near $z$ where a parallel copy of $G$ intersects $F$, and the resulting cap $c_G$ for $F$ is dual to both $c_W$ and $c_{W'}$.

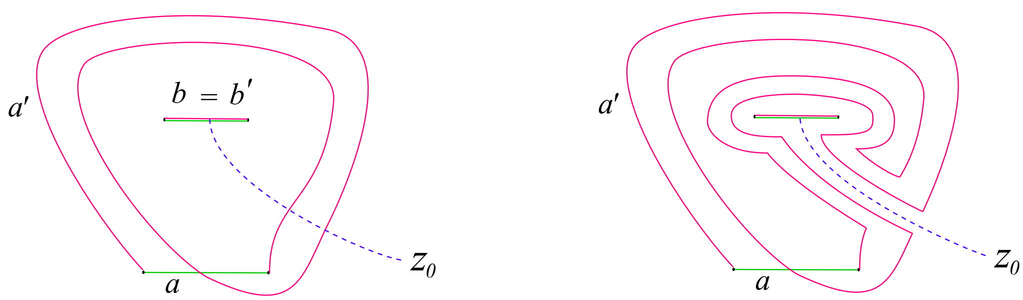


FIGURE 11. The case of one pair of Whitney disks, with $f(z_0) = z = f \cap G$. Slides are done in the order starting closest to $f(b) = \beta = \beta' = f(b')$.

Now consider the case that $r$ and $z = G \cap F$ do not lie in the same connected component of $F \smallsetminus (\partial c_W \cup \partial c_{W'})$, and observe that this means that $\beta$ and $z$ do not lie in the same component of the complement in $f$ of the immersed loop $\alpha \cup \alpha'$ (see the left side of Figure 11). In this case we can modify the original Whitney disk $W'$ before constructing $F$ using the sliding maneuver of Section 2.D to arrange that $\beta$ and $z$ do lie in the same component of $f \smallsetminus (\alpha \cup \alpha')$:

Since $S^2 \smallsetminus \partial W$ is connected, there is an embedded arc $\gamma$ from $z$ to $r \in \beta' = \beta$ such that $\gamma$ is disjoint from $\alpha$ (the preimage of $\gamma$ is the dashed blue arc in Figure 11). Eliminate the intersections between $\gamma$ and $\alpha'$ by sliding $W'$ over itself from $\alpha'$ to $\beta'$ guided by $\gamma$ as in

Section 2.D (right side of Figure 11). By Lemma 2.10 this does not change the isotopy class of $f_{W'}$, and now the construction of the cap $c_G$ for $F$ goes through as desired.

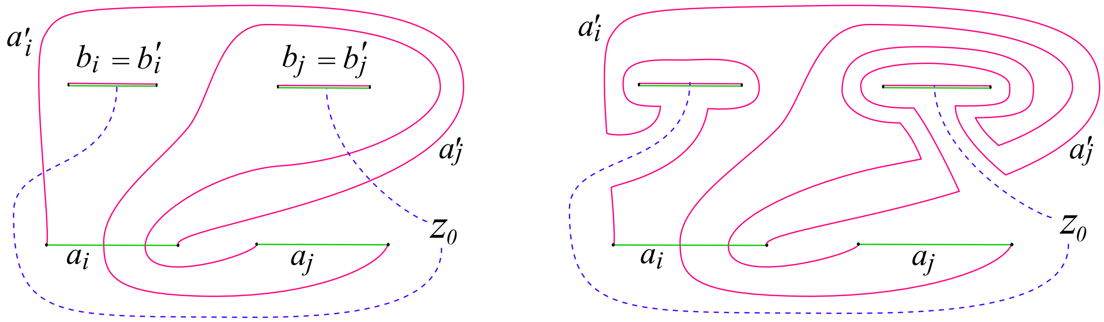


FIGURE 12. The general multiple Whitney disk case of Figure 11. Again the arcs $\gamma_i$ lead to our slides algorithmically, starting closest to the $\beta$-arcs.

For the general statement, apply the same construction to each of the pairs of Whitney disks $W_i$ and $W_i'$ in $\mathcal{W}$ and $\mathcal{W}'$. Start with disjointly embedded arcs $\gamma_i$ in $S^2 \smallsetminus \partial\mathcal{W}$ from the common arcs $\beta_i$ to $z$. The only new complication is that making these arcs disjoint from $\partial\mathcal{W}'$ may involve more Whitney disk slides as shown in Figure 12. By Lemma 2.2 and Lemma 2.10 these sides preserve the isotopy class of $f_{\mathcal{W}'}$, and by applying Lemma 2.7 to each pair $W_i, W_i'$ we have that $f_{\mathcal{W}}$ is isotopic to $f_{\mathcal{W}'}$. $\square$

## 3. New Proof of Gabai's LBT

Let $M$ be a smooth orientable 4–manifold and $f : S^2 \looparrowright M$ a generic smooth map with $0 = \mu(f) \in \mathbb{Z}[\pi_1 M]/\langle g - g^{-1}\rangle$ and with geometric dual $G$. Recall that $\mathcal{R}^G_{[f]}$ denotes the set of isotopy classes of embedded spheres which are homotopic to $f$ and have $G$ as a geometric dual.

**Outline of our proof of Gabai's LBT:** We will show that $\mathcal{R}^G_{[f]}$ contains a unique element if $\pi_1 M$ does not contain 2-torsion: As explained in Section 2.A, any two embedded spheres in $\mathcal{R}^G_{[f]}$ are related via a finite sequence of isotopies, finger moves and then Whitney moves, all away from $G$. Denoting the result of the finger moves by $f$, we will consider all possible collections of Whitney disks on $f$ in $M \smallsetminus G$ and show that all the resulting embeddings are isotopic. As a first step, Section 3.A describes precisely the various types of choices involved in constructing a collection $\mathcal{W}$ of clean Whitney disks on $f$ such the result $f_{\mathcal{W}}$ of doing the Whitney moves in $\mathcal{W}$ on $f$ is an embedding. After all, any such $\mathcal{W}$ corresponds to choices of local sheets at self-intersections, pairings of self-intersections, Whitney arcs and finally Whitney disks. In Sections 3.B–3.G we show that such choices always exist and prove that the isotopy class of $f_{\mathcal{W}}$ does not depend on any of these choices.

### 3.A. Choices of sheets, pairings, W-arcs and W-disks.

We'll discuss the four types of choices $\mathsf{C}_{\text{sheets}}$, $\mathsf{C}_{\text{pairings}}$, $\mathsf{C}_{\text{W-arcs}}$ and $\mathsf{C}_{\text{W-disks}}$ that determine a clean collection $\mathcal{W}$ of Whitney disks on $f : S^2 \looparrowright M$ and hence a generic homotopy from $f$ to an embedding $f_{\mathcal{W}}$ (with geometric dual $G$). In the following, each step will depend on having made all previous choices. Moreover, each later choice lets us reconstruct the previous choices.

Denote the set of transverse self-intersections of $f$ by $\{p_1, \ldots, p_{2n}\} \subset f(S^2)$, where the ordering of the $p_i$ is an artifact of the notation and will never be used; and fix a whisker for $f$ from the basepoint of $M$.

$\mathsf{C}_{\text{sheets}}$: A *choice of sheets* $\{x_1, \ldots, x_{2n}\} \in \mathsf{C}_{\text{sheets}}$ consists of choices $x_i \in f^{-1}(p_i) \subset S^2$, subject to the following requirement: By Section 2.C, each $x_i$ orients a double point loop at $p_i$ by the convention that the loop is the image of a path *starting from* $x_i$. Via the whisker for $f$ we get a well-defined group element $g(x_i) \in \pi_1 M$.

Then our choice of sheets is required to satisfy

$$(*) \qquad 0 = \sum_{i=1}^{2n} \epsilon_i \cdot g(x_i) \in \mathbb{Z}[\pi_1 M], \text{ where } \epsilon_i \in \{\pm 1\} \text{ is the sign of } p_i.$$

A different choice of whisker for $f$ would change each $g(x_i)$ to a conjugate $g(x_i)^g$ for some fixed $g \in \pi_1 M$, hence our requirement $(*)$ is independent of the whisker. Moreover, switching the preimage choice $x_i$ at $p_i$ has the effect of inverting the group element $g(x_i)$, so choices of sheets exist since $0 = \mu(f) \in \mathbb{Z}[\pi_1 M]/\langle g - g^{-1}\rangle$.

$\mathsf{C}_{\text{pairings}}$: For $\{x_1, \ldots, x_{2n}\} \in \mathsf{C}_{\text{sheets}}$, a *compatible choice of pairings* $\{x_1^\pm, \ldots, x_n^\pm\} \in \mathsf{C}_{\text{pairings}}$ consists of $n$ distinct pairs $x_i^\pm := (x_i^+, x_i^-) = (x_{j_i}, x_{k_i})$ with $\epsilon_{j_i} = +1 = -\epsilon_{k_i}$ and $g(x_{j_i}) = g(x_{k_i})$. A choice of pairings exists by our requirement $(*)$ on $\{x_1, \ldots, x_{2n}\}$ and it induces pairings $(p_i^+, p_i^-)$ of the self-intersections of $f$.

$\mathsf{C}_{\text{W-arcs}}$: For $\{x_1^\pm, \ldots, x_n^\pm\} \in \mathsf{C}_{\text{pairings}}$, a *compatible choice of Whitney arcs* $\{\alpha_1, \beta_1, \ldots, \alpha_n, \beta_n\} \in \mathsf{C}_{\text{W-arcs}}$ are the images under $f$ of disjointly embedded arcs $a_i \subset S^2$ joining $x_i^+$ and $x_i^-$, and arcs $b_i \subset S^2$ joining $y_i^+$ and $y_i^-$ for $i = 1, \ldots, n$, where $f^{-1}(p_k^\pm) = \{x_k^\pm, y_k^\pm\}$. Here $\alpha_i := f(a_i)$ and $\beta_i := f(b_i)$ are disjoint, except that $\partial\alpha_i = \{p_i^+, p_i^-\} = \partial\beta_i$. Note that $\alpha_i \subset f(S^2)$ determines $a_i \subset S^2$ and hence the original choice of pairings is determined by $\{\alpha_1, \ldots, \alpha_n\}$ alone.

$\mathsf{C}_{\text{W-disks}}$: Given a choice of Whitney arcs $\{\alpha_1, \beta_1, \ldots, \alpha_n, \beta_n\} \in \mathsf{C}_{\text{W-arcs}}$, a *compatible choice of Whitney disks* $\{W_1, \ldots, W_n\} \in \mathsf{C}_{\text{W-disks}}$ is a clean collection of Whitney disks $W_i$ whose boundaries are equal to the circles $\alpha_i \cup \beta_i \subset M$. Recall that *clean* means the $W_i$ are framed, disjointly embedded, have interiors disjoint from $f$, and are disjoint from $G$. The existence of a choice of Whitney disks for any choice of Whitney arcs follows from Lemma 2.4. To reconstruct $\alpha_i$ from $W_i : D^2 \hookrightarrow M$, we also require that $\alpha_i = W_i(S^1_-)$, where $S^1_- \subset S^1 = \partial D^2 \subset D^2 \subset \mathbb{R}^2$ is the lower semi-circle.

In the following, we will abbreviate our choices by

$$\mathsf{x} := \{x_1, \ldots, x_{2n}\},\ \mathsf{x}^\pm := \{x_1^\pm, \ldots, x_n^\pm\},\ \mathcal{A} := \{\alpha_1, \beta_1, \ldots, \alpha_n, \beta_n\} \text{ and } \mathcal{W} := \{W_1, \ldots, W_n\}.$$

The meaning of $\partial\mathcal{W} = \mathcal{A}$ should be clear from our conventions. The embedded sphere obtained from $f$ by doing Whitney moves guided by the Whitney disks in $\mathcal{W}$ is denoted $f_{\mathcal{W}}$.

3.B. **Existence and choices of Whitney disks.** For future reference we observe here that the existence of a compatible $\mathcal{W} \in \mathsf{C}_{\text{W-disks}}$ for any given $\mathcal{A} \in \mathsf{C}_{\text{W-arcs}}$ guaranteed by Lemma 2.4, together with the definitions of pairing choices and sheet choices in section 3.A, imply the following:

**Lemma 3.1.** *Given* $\mathsf{x}^\pm \in \mathsf{C}_{\textit{pairings}}$*, there exists* $\mathcal{W} \in \mathsf{C}_{\textit{W-disks}}$ *compatible with* $\mathsf{x}^\pm$*.*
*And given* $\mathsf{x} \in \mathsf{C}_{\textit{sheets}}$*, there exists* $\mathcal{W} \in \mathsf{C}_{\textit{W-disks}}$ *compatible with* $\mathsf{x}$*.* $\square$

From Lemma 2.8, the isotopy class of $f_{\mathcal{W}}$ is independent of the interiors of the Whitney disks in $\mathcal{W}$, i.e. $f_{\mathcal{W}}$ only depends on $\mathcal{A}$.

We next introduce Norman spheres, which will play a key role in showing that the isotopy class of $f_{\mathcal{W}}$ is also independent of choices of arcs and pairings for any given sheet choice.

3.C. **Norman spheres.** Fix a choice of sheets $\mathsf{x} = \{x_1, x_2, \ldots, x_{2n}\} \in \mathsf{C}_{\text{sheets}}$ for $f$. We need yet another type of choice to define a Norman sphere (whose isotopy class will ultimately only depend on $\mathsf{x}$). Recall that $D_z \subset f$ denotes a small disk around $z = f \cap G$ such that each point in $D_z$ intersects a parallel of $G$ which is geometrically dual to $f$.

$\mathsf{C}_{\text{N-arcs}}$: A *compatible choice of Norman arcs* $\mathcal{Z} := \{\sigma_1, \ldots, \sigma_{2n}\} \in \mathsf{C}_{\text{N-arcs}}$ for $\mathsf{x} \in \mathsf{C}_{\text{sheets}}$ is the image under $f$ of disjointly embedded arcs $s_i \subset S^2$ starting at $x_i$ and ending in $f^{-1}(\partial D_z)$. Then $\sigma_i := f(s_i) \subset f(S^2)$ are disjointly embedded arcs starting at $p_i$ and ending in $\partial D_z$; they determine the arcs $s_i$ uniquely.

**Definition 3.2.** The *Norman sphere* $f^G_{\mathcal{Z}} : S^2 \hookrightarrow M$ is obtained from $f$, $G$ and $\mathcal{Z}$ by eliminating all the self-intersections $p_i \in f \pitchfork f$ by tubing $f$ into parallel copies of $G$ along the $\sigma_i$. Precisely, these tubing operations replace the image of a small disk around each $y_i \in S^2$ by a normal tube along $\sigma_i$ together with a parallel copy $G_i$ of $G$ with a small normal disk to $f$ removed at $G_i \cap f$. Here $f^{-1}(p_i) = \{x_i, y_i\}$ with $x_i \in \mathsf{x}$, and the $y_i$-sheet of $f$ at $p_i$ is deleted by the tubing operation since the $y_i$-sheet is normal to $\sigma_i$ at $p_i$.

By construction, the Norman sphere $f^G_{\mathcal{Z}}$ is embedded and has $G$ as a geometric dual. Also, $f^G_{\mathcal{Z}}$ is homotopic to $f$ since the copies of $G$ in the connected sum with $f$ come in oppositely oriented pairs having the same group element by our requirement $(*)$ in Section 3.A on the sheet choice $\mathsf{x}$. Hence $f^G_{\mathcal{Z}} \in \mathcal{R}^G_{[f]}$.

Surprisingly, we will show in Lemma 3.5 that the isotopy class of $f^G_{\mathcal{Z}}$ only depends on $\mathsf{x}$ and not at all on $\mathcal{Z}$.

We remark that the $\sigma_i$ are as in [3] which are the simplest of the three types of arcs used by Gabai. The $\sigma_i$ in [3] are allowed to intersect but here we require them to be disjointly embedded.

**Lemma 3.3.** *For any given choice of sheets* $\mathsf{x}$*, if* $\mathcal{W}$ *is an* $\mathsf{x}$*-compatible choice of Whitney disks then there is an* $\mathsf{x}$*-compatible choice of Norman arcs* $\mathcal{Z}$ *such that* $f^G_{\mathcal{Z}}$ *is isotopic to* $f_{\mathcal{W}}$*.*

*Proof.* We apply the first step in the proof of Lemma 2.9 simultaneously to all $\alpha_i$: Let $\mathcal{A} := \partial\mathcal{W}$ be the Whitney arcs and $\mathsf{x}^{\pm}$ be the choice of pairings determined by $\mathcal{A}$. To construct the Norman arcs $\mathcal{Z}$, isotope the Whitney arcs $\alpha_i$ just across $\partial D_z$ and extend this isotopy to an isotopy of $W_i$ in a collar on $\alpha_i$; see Figure 13 where $D_{z_0} := f^{-1}(D_z)$. This can be done keeping the $\alpha_i$ disjoint from each other and from all $\beta_j$. Deleting the part of the new $\alpha_i$ that lies in the interior of $D_z$ gives two arcs $\sigma_i^{\pm}$ which start at $x_i^{\pm}$ and end in $\partial D_z$.

Define $\mathcal{Z} := \{\sigma_1^-, \sigma_1^+, \ldots, \sigma_n^-, \sigma_n^+\}$ and observe that since the corresponding copies $G_i^{\pm}$ of $G$ are oppositely oriented, the Norman sphere $f^G_{\mathcal{Z}}$ is isotopic to the result of tubing $f$ to itself along each $\alpha_i \subset \partial W_i$, then surgering a meridional cap dual to $\alpha_i$ that has been tubed into $G$ as in Figure 8. So $f^G_{\mathcal{Z}}$ is isotopic to $f_{\mathcal{W}}$ by Lemma 2.7. $\square$

In the proofs of the next two lemmas we describe isotopies of Norman spheres using homotopies of Norman arcs by requiring that the radii of the tubes are not equal at any temporarily-created intersection between Norman arcs during a homotopy. Following Gabai,

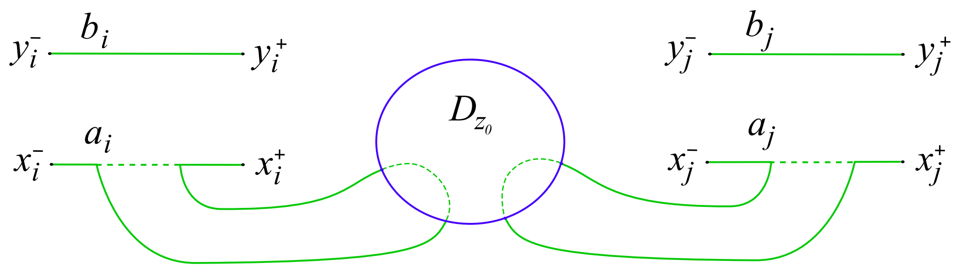


FIGURE 13. The preimages $a_i$ and $a_j$ of arcs $\alpha_i$ and $\alpha_j$ after the isotopy.

we indicate the tube of smaller radius as an under-crossing of the corresponding Norman arc.

**Lemma 3.4** (Lemma 5.11(ii) of [3])**.** *Given any $\mathcal{Z}' \in \mathsf{C}_{N\text{-}arcs}$ and points $z_1, \dots, z_{2n} \in \partial D_z$, there is a choice of Norman arcs $\mathcal{Z} = \{\sigma_1, \dots, \sigma_{2n}\}$, compatible with the same $\mathsf{x} \in \mathsf{C}_{sheets}$ as $\mathcal{Z}'$, such that $\sigma_i$ ends in $z_i$ and the Norman spheres $f^G_{\mathcal{Z}'}$ and $f^G_{\mathcal{Z}}$ are isotopic.*

*Proof.* It suffices to observe that neighboring $z_i$ and $z_j$ in $\partial D_z$ can be exchanged by pushing the tube around $\sigma_j$ across (and inside) the tube around $\sigma_i$, as in Figure 14 and Figure 15. $\square$

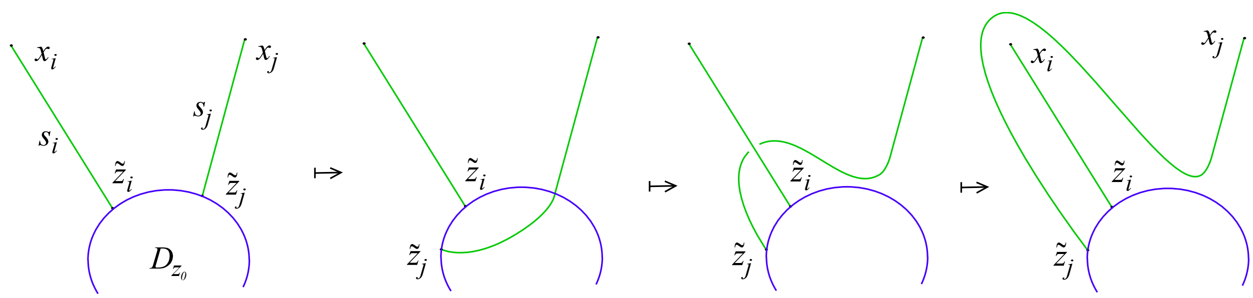


FIGURE 14. The indicated homotopy of $s_i$ and $s_j$ corresponds to an isotopy of Norman spheres which slides the tube around $\sigma_j$ inside of the tube around $\sigma_i$. See Figure 15.

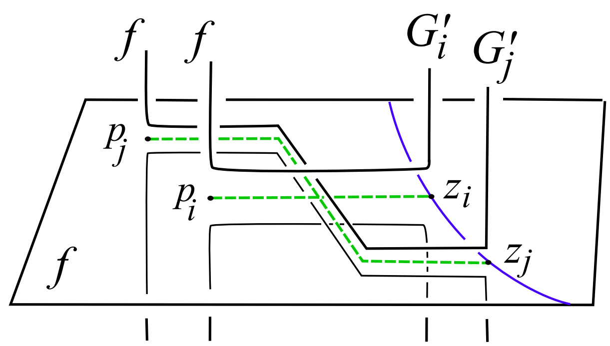


FIGURE 15. The image of the third-from-left picture in Figure 14. Here the smaller radius of the tube around $\sigma_j$ compared to the tube around $\alpha_i$ corresponds to $s_j$ crossing under $s_i$ in Figure 14.

**Lemma 3.5.** *If two choices of Norman arcs $\mathcal{Z}, \mathcal{Z}' \in \mathsf{C}_{\text{N-arcs}}$ are compatible with the same $\mathsf{x} \in \mathsf{C}_{\text{sheets}}$ then the Norman spheres $f^G_{\mathcal{Z}}$ and $f^G_{\mathcal{Z}'}$ are isotopic.*

As a consequence, we get a Norman sphere $f^G_{\mathcal{Z}} =: f^G_{\mathsf{x}} \in \mathcal{R}^G_{[f]}$ for a given choice of sheets $\mathsf{x}$.

*Proof.* Let $\mathsf{x}^{\pm}$ be any compatible choice of pairings for $\mathsf{x}$. By Lemma 3.4 we may assume that $\mathcal{Z} = \{\sigma_1^-, \sigma_1^+, \dots, \sigma_n^-, \sigma_n^+\}$ induces the cyclic ordering $(z_1^-, z_1^+, z_2^-, z_2^+, \dots, z_n^-, z_n^+)$ in $\partial D_z$, where $z_i^{\pm}$ is the end-point of $\sigma_i^{\pm}$.

We will first construct a choice of Whitney disks $\mathcal{W}$ for $f$ such that $f^G_{\mathcal{Z}}$ is isotopic to $f_{\mathcal{W}}$, by performing essentially the inverse of the steps in the proof of Lemma 3.3: For each $i$, denote by $\alpha_i$ the union of the embedded arcs $\sigma_i^-$ and $\sigma_i^+$ together with a short arc in $\partial D_z$ that runs between $z_i^+$ and $z_i^-$. These $\alpha_i$ then form one half of a collection of Whitney arcs for the choice of pairings $\mathsf{x}^{\pm}$. By choosing appropriate $\beta_i$ we get a $\mathsf{x}^{\pm}$-compatible choice of Whitney arcs $\mathcal{A} = \{\alpha_1, \beta_1, \dots, \alpha_n, \beta_n\}$, and by Lemma 2.4 there exists a collection $\mathcal{W} \in \mathsf{C}_{\text{W-disks}}$ with boundary $\mathcal{A}$.

It follows that $f^G_{\mathcal{Z}}$ is isotopic to $f_{\mathcal{W}}$ by Lemma 2.7, since $f^G_{\mathcal{Z}}$ is isotopic to the result of surgering the capped surface formed by tubing $f$ along the $\alpha_i$ arcs, as observed in the proof of Lemma 3.3.

Applying the first part of this construction to the Norman arcs $\mathcal{Z}'$ yields a half collection of disjointly embedded Whitney arcs $\alpha'_i$ formed from $\sigma'^{\pm}_i$ by adding short arcs in $\partial D_z$.

Now pause to observe that if these $\alpha'_i$ are each disjoint from all the previously-chosen $\beta_j$, then the unions $\alpha'_i \cup \beta_i$ are Whitney circles for a clean collection $\mathcal{W}'$ of Whitney disks on $f$ by Lemma 2.4, and $f^G_{\mathcal{Z}}$ is isotopic to $f^G_{\mathcal{Z}'}$: The collections $\mathcal{W}$ and $\mathcal{W}'$ share the common $\beta_i$-arcs so $f_{\mathcal{W}}$ is isotopic to $f_{\mathcal{W}'}$ by Proposition 2.11. And analogously to the above argument, we see that $f^G_{\mathcal{Z}'}$ is isotopic to $f_{\mathcal{W}'}$, completing the proof.

So it just remains to get $\alpha'_i \cap \beta_j = \emptyset$ for all $i, j$.

Since the $\alpha'_i$ are constructed from $\sigma'^{\pm}_i$ by adding short arcs in $\partial D_z$, it suffices to show that we may push all the $\sigma'^{\pm}_i$ off of all the $\beta_j$ in a way that corresponds to an isotopy of the Norman sphere $f^G_{\mathcal{Z}'}$. It will be convenient to describe this pushing-off construction in the domain of $f$, so we want to get $s'^{\pm}_i \cap b_j = \emptyset$, where $b_j \subset S^2$ is an arc from $y_j^-$ to $y_j^+$ with $f(b_j) = \beta_j$, and $s'^{\pm}_i \subset S^2$ goes from $x_i^{\pm}$ to $f^{-1}(z_i^{\pm})$ with $f(s'^{\pm}_i) = \sigma'^{\pm}_i$.

Our construction will work with one $b_j$ at a time, removing intersections with all $s'^{\pm}_i$ in a way that does not create new intersections in any previously cleaned-up $b_k$. This will be accomplished by describing an isotopy of the Norman sphere tubes induced by pushing (as needed) each $s'^{\pm}_i$ across the endpoints $y_j^{\pm}$ of $b_j$, using the fact that a disk around $y_j^{\pm}$ maps to a disk in the Norman sphere consisting of a tube along $\sigma_j^{\pm}$ into $G_j^{\pm}$. As observed by Gabai [3, Rem.5.10], in the case $i = j$ we are *not* able to push $s'^{\pm}_j$ across $y_j^{\pm}$, but we *are* able to push $s'^{\pm}_j$ across the opposite-signed $y_j^{\mp}$. This is similar to the fact that a handle cannot be slid over itself.

Consider first the case where some $b_j$ only has intersections with a single $s'^{\pm}_i$ (Figure 16 left). If $i \neq j$ then these intersections can all be eliminated by an isotopy of $s'^{\pm}_i$ across $y_j^+$ (Figure 16 right). If $i = j$ then $b_j \cap s'^{\pm}_j$ can be eliminated by an isotopy of $s'^{\pm}_j$ across the oppositely-signed $y_j^{\mp}$. These isotopies pushing $s'^{\pm}_j$ off of $b_j$ can be done without creating any intersections among the parallel strands of $s'^{\pm}_j$.

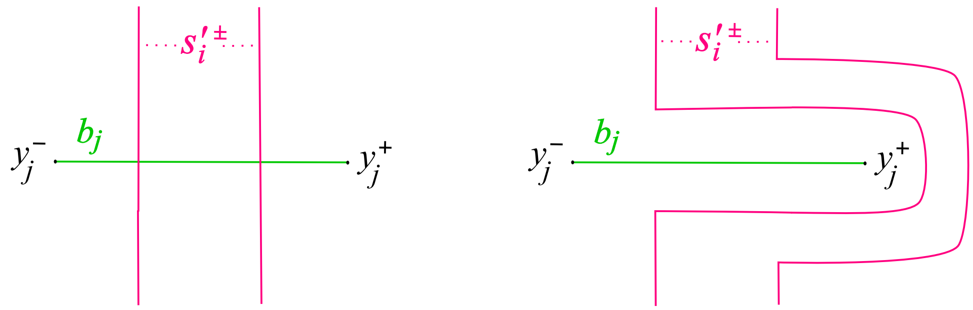


FIGURE 16. For $i \neq j$ all strands of $s_i'^{\pm}$ can be pushed off $b_j$ across $y_j^+$.

Next consider the case where $b_j$ intersects only the two arcs $s_j'^{+}$ and $s_j'^{-}$, each in a single point $r^+ = b_j \cap s_j'^{+}$ and $r^- = b_j \cap s_j'^{-}$. If $r^{\pm}$ is adjacent to $y_j^{\mp}$ in $b_j$, then each $r^{\pm}$ can be eliminated as in the previous case by pushing $s_j'^{\pm}$ across $y_j^{\mp}$. If $r^{\pm}$ is adjacent to $y_j^{\pm}$ in $b_j$, then first eliminate $r^-$ by pushing $s_j'^{-}$ across $y_j^+$ and under $s_j'^{+}$, as in Figure 17 left. Then eliminate $r^+$ by pushing $s_j'^{+}$ across $y_j^-$ and over $s_j'^{-}$, as in Figure 17 center. At this point we have $b_j \cap s_j'^{\pm} = \emptyset$, but $s_j'^{+}$ intersects $s_j'^{-}$ in two points $q$ and $q'$. Each of $q$ and $q'$ can be eliminated by pushing $s_j'^{-}$ along $s_j'^{+}$ and across (under) $x_j^+$ as in Figure 17 right, since the tube around $\sigma_j^-$ has a smaller radius. Note that the pushing of $s_j'^{-}$ along $s_j'^{+}$ will create new intersections between $s_j'^{-}$ and any other $b_k$ with $k \neq j$ that intersected $s_j'^{+}$ along the strand of the original $s_j'^{+}$ between $x_j^+$ and $r^+$. But such new intersections only are created in a $b_k$ that has yet to be cleaned up.

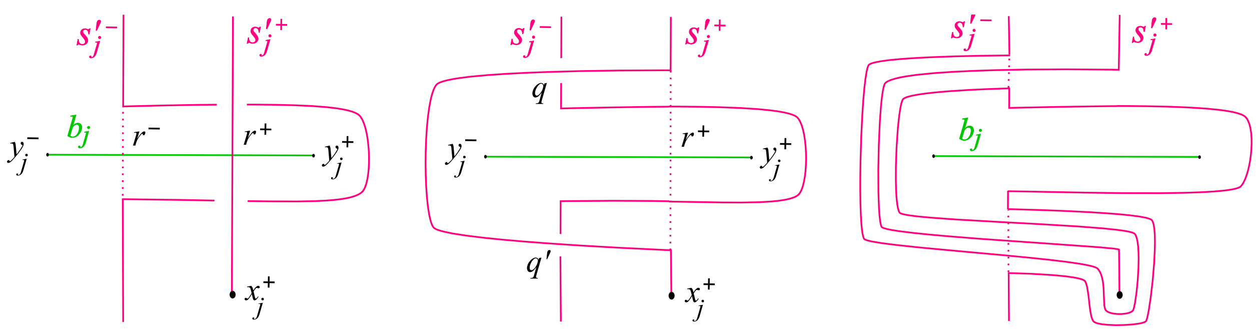


FIGURE 17

The construction of the previous paragraph can be adapted to the general case where $b_j$ intersects arbitrary strands of $s_i'^{\pm}$ for arbitrary $i$ as follows. (Picture the $s_j'^{\pm}$-arcs in Figure 17 as two among several parallel collections of strands.) First simultaneously push all strands of $s_j'^{-}$ and all strands of any other $s_i'^{\pm}$ with $i \neq j$ under any and all strands of $s_j'^{+}$ and across $y_j^+$. This can be done in parallel, without creating any intersections among the strands that are being isotoped. Then simultaneously push any and all strands of $s_i'^{+}$ over all other strands and across $y_j^-$. This can be done in parallel, so that the only resulting intersections between $s'$-arcs are where $s_i'^{+}$ passes over other strands. At this point $b_j$ is disjoint from all $s_i'^{\pm}$, and the intersections among $s'$-arcs can all be eliminated by pushing the under-crossing arcs along $s_j'^{+}$ across (under) $x_j^+$. □

3.D. **Independence of pairings and Whitney arcs.** From Lemmas 3.3 and 3.5 we get:

**Corollary 3.6.** *If two choices of Whitney disks $\mathcal{W}, \mathcal{W}' \in \mathsf{C}_{\textit{W-disks}}$ are each compatible with the same choice of sheets $\mathsf{x} \in \mathsf{C}_{\textit{sheets}}$, then $f_{\mathcal{W}}$ is isotopic to $f_{\mathcal{W}'}$. In particular, $f_{\mathcal{W}} \in \mathcal{R}^G_{[f]}$ is independent of* $\mathsf{x}$*-compatible choices of pairings, Whitney arcs and Whitney disks.* $\square$

As a consequence, $f_{\mathcal{W}} \in \mathcal{R}^G_{[f]}$ only depends on $\mathsf{x}$ and it's safe to write $f_{\mathcal{W}} =: f_{\mathsf{x}} \in \mathcal{R}^G_{[f]}$, where the existence of an $\mathsf{x}$-compatible $\mathcal{W}$ is guaranteed by Lemma 3.1.

By the same lemmas we also see that $f_{\mathsf{x}}$ is isotopic to the Norman sphere $f^G_{\mathsf{x}}$, whose isotopy class therefore only depends on $\mathsf{x}$ but not on $G$.

To complete the proof of Gabai's LBT it remains to consider the $\mathsf{x}$-dependence of $f_{\mathsf{x}}$.

3.E. **Double sheet changes.** Let $\mathsf{x} = \{x_1, \ldots, x_{2n}\} \in \mathsf{C}_{\text{sheets}}$ and recall that $g(x_i) \in \pi_1 M$ is represented by a double point loop through $p_i$ which is the image of an oriented arc from $x_i$ to $y_i$, where $f^{-1}(p_i) = \{x_i, y_i\}$. Switching the choice $x_i$ to $y_i$ changes $g(x_i)$ to $g(y_i) = g(x_i)^{-1}$ while keeping the sign $\epsilon_i$ of $p_i$. Changing the whisker for $f$ changes all $g(x_i)$ by a fixed conjugation and also keeps the signs.

Assume that for two indices $i, j$ we have $\epsilon_j = -\epsilon_i$ and $g(x_j) = g(x_i) =: g$. Then a different choice of sheets $\mathsf{x}' \in \mathsf{C}_{\text{sheets}}$ can be defined by replacing $x_i$ by $y_i$ and replacing $x_j$ by $y_j$, since it satisfies our requirement $(*)$ in Section 3.A with the canceling terms $\epsilon_i \cdot g + \epsilon_j \cdot g = 0$ replaced by $\epsilon_i \cdot g^{-1} + \epsilon_j \cdot g^{-1} = 0$.

We will refer to such a change of sheet choice as a *double sheet change.*

**Lemma 3.7.** *If* $\mathsf{x}, \mathsf{x}' \in \mathsf{C}_{\textit{sheets}}$ *differ by a double sheet change, then* $f_{\mathsf{x}} = f_{\mathsf{x}'} \in \mathcal{R}^G_{[f]}$.

*Proof.* Let $\{x_i, x_j\} \subset \mathsf{x}$ be the local sheets involved in the double sheet change. There is a choice of pairings $\mathsf{x}^{\pm}$ compatible with $\mathsf{x}$ such that $x_i = x_1^+$ and $x_j = x_1^-$ (or vice versa). Moreover, by Lemma 3.1 there is a choice of Whitney disks $\mathcal{W} = \{W_1, \ldots, W_n\}$ compatible with $\mathsf{x}^{\pm}$, i.e. $p_i$ and $p_j$ are paired by $W_1$.

Let $\mathcal{W}' := \{W_1', W_2, \ldots, W_n\}$ be the choice of Whitney disks where $W_1'$ differs from $W_1$ only by precomposing with a reflection of the domain $D^2$ across the horizontal diameter. This exchanges the two boundary arcs of $W_1$ but does not change the effect of doing a Whitney move since $W_1$ and $W_1'$ have the same image in $M$. Now observe that $\mathcal{W}'$ is compatible with $\mathsf{x}'$ and it follows from Corollary 3.6 that $f_{\mathsf{x}} = f_{\mathcal{W}} = f_{\mathcal{W}'} = f_{\mathsf{x}'} \in \mathcal{R}^G_{[f]}$. $\square$

3.F. **Choice of sheets for double point loops not of order 2.** Consider a sheet choice $\mathsf{x} = \{x_1, \ldots, x_{2n}\} \in \mathsf{C}_{\text{sheets}}$ such that for some $i$ we have $g_i := g(x_i) \in \pi_1 M$ with $g_i^2 \neq 1$. If $\mathsf{x}'$ is a different choice of sheets that takes $y_i$ as the preferred preimage instead of $x_i$, then this has the effect of inverting $g_i$. Since $g_i \neq g_i^{-1}$, in order for $\mathsf{x}'$ to satisfy the requirement $(*)$ in Section 3.A of a choice of sheets it follows that $\mathsf{x}'$ must also switch some oppositely-signed $x_j$ to $y_j$, where $g(x_j) = g(x_i)$. So $\mathsf{x}'$ differs from $\mathsf{x}$ by at least one double sheet change, and Lemma 3.7 applied finitely many times gives:

**Lemma 3.8.** *If choices of sheets* $\mathsf{x}, \mathsf{x}' \in \mathsf{C}_{\textit{sheets}}$ *only differ at self-intersections $p_i$ where the double point loops $g_i$ satisfy $g_i^2 \neq 1$, then $f_{\mathsf{x}} = f_{\mathsf{x}'} \in \mathcal{R}^G_{[f]}$.* $\square$

Note that the assumption does not depend on the whisker for $f$.

3.G. **Choice of sheets for trivial double point loops.** Let $p_i$ be a self-intersection of $f$ with trivial group element $1 \in \pi_1 M$. By the same construction as in the proof of Lemma 2.4, $p_i$ admits a clean *accessory disk* $A_i$, i.e. $A_i$ is a framed embedded disk with interior disjoint from $f$ such that the boundary circle $\partial A_i \subset f$ changes sheets just at $p_i$. See [7, Sec.7] for details on accessory disks. If $p_i^+$ and $p_i^-$ are oppositely-signed with trivial group element, then clean Whitney disks for $p_i^\pm$ can be constructed by banding together two clean accessory disks $A_i^\pm$ as in Figure 18, which shows two choices of bands resulting in Whitney disks $W_i$ and $W_i'$ which induce the possible different sheet-choices. These Whitney disks are supported in a neighborhood of the union of the two accessory disks together with a generic disk in $f$ containing the accessory circles $\partial A_i^\pm$. We will show that $W_i$ and $W_i'$ are isotopic via an ambient isotopy supported near one of the accessory disks. Hence $f_{W_i}$ is isotopic to $f_{W_i'}$.

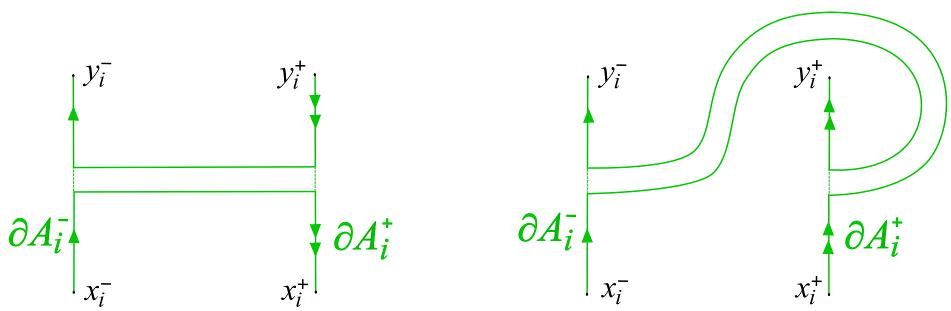


FIGURE 18. Preimages of Whitney circles for $W_i$ (left) and $W_i'$ (right) formed by banding together accessory disks $A_i^\pm$ in two different ways, with $W_i$ satisfying the sheet choice $\{x_i^-, x_i^+\}$ and $W_i'$ satisfying the sheet choice $\{x_i^-, y_i^+\}$. Applying the rotation isotopy of Lemma 3.9 to $A_i^+$ interchanges $x_i^+$ and $y_i^+$.

A regular neighborhood of a clean accessory disk is diffeomorphic to a standard model in 4–space, so we work locally, dropping superscripts and subscripts:

Let $(\Delta, \partial\Delta) \looparrowright (B^4, S^3)$ be a generic 2–disk with a single self-intersection $p$ which is the result of applying a cusp-homotopy [2, 1.6] to a standard $(D^2, S^1) \subset (B^4, S^3)$. Then $p$ admits a clean accessory disk $A$, and the following lemma will be proved:

**Lemma 3.9.** *There is an ambient isotopy $h_s$ of $B^4$ such that*

(1) *$h_0$ is the identity,*
(2) *$h_1(\Delta \cup A) = \Delta \cup A$, and*
(3) *$h_1|_{\partial\Delta}$ is rotation by 180 degrees, inducing a reflection of $\partial A$.*

Applying Lemma 3.9 to a $B^4$-neighborhood of $A_i^+$ we see that the two Whitney disks $W_i$ and $W_i'$ in Figure 18 are isotopic: Rotating the right accessory arc $\partial A_i^+$ by 180 degrees drags one band to the other, and hence one Whitney disk to the other.

*Proof.* To prove Lemma 3.9, consider $\Delta$ as the trace of a null-homotopy of the Whitehead double of the unknot in $S^3 = \partial B^4$ which pulls apart the clasp in a collar $S^3 \times I \subset B^4$, creating the self-intersection $p$ admitting a clean accessory disk $A$, as in Figure 19. Define the homotopy $h_s$ of $\Delta$ in the coordinates of Figure 19 to be rotation around the horizontal by $180s$ degrees in each $S^3$-slice of $S^3 \times I$ and the identity on $I$. Extend to $B^4$ by tapering the rotation back to zero inside the collar. $\square$

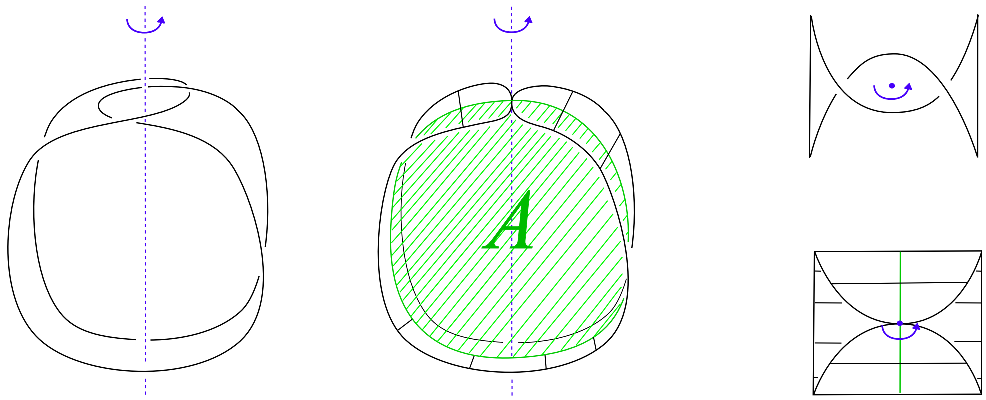


FIGURE 19. Left: The Whitehead double of the unknot in $S^3$ is the boundary of $\Delta$. Center: The clean accessory disk $A$ for the self-intersection $p$ of $\Delta$ which corresponds to the clasp singularity. Both $\Delta$ and $A$ have 180 degree rotational symmetry (top views of left and center on upper and lower right).

By Corollary 3.6 we can compute $f_{\mathsf{x}} = f_{\mathcal{W}}$ by $\mathcal{W} \in \mathsf{C}_{\text{W-disks}}$ whose Whitney disks pairing self-intersections with trivial group elements are formed from banding together accessory disks as above. So in combination with Lemma 3.8 we have:

**Corollary 3.10.** *If choices of sheets $\mathsf{x}, \mathsf{x}'$ only differ at self-intersections $p_i$ whose double point loops don't have order* 2*, then $f_{\mathsf{x}} = f_{\mathsf{x}'} \in \mathcal{R}^G_{[f]}$.*

This result completes the proof of Gabai's LBT. To prove our main Theorem 1.1 it remains to understand the $\mathsf{x}$-dependence of $f_{\mathsf{x}}$ in the presence of self-intersections with group elements of order 2. In the subsequent Section 4 and Section 5 we will show that it is completely controlled by the Freedman–Quinn invariant.

## 4. The Freedman–Quinn invariant

In Section 4.A we review some relevant aspects of the intersection form on $\pi_3$ of a 6–manifold. In Section 4.B the Freedman–Quinn invariant is defined using the self-intersection invariant applied to the track of a homotopy between spheres in $M^4 \times \mathbb{R}$, which is a map of a 3–manifold to a 6–manifold rel boundary.

4.A. **3–manifolds in 6–manifolds.** Recall that for a smooth oriented 6–manifold $P^6$, the intersection and self-intersection invariants give maps

$$\lambda_3 : \pi_3 P \times \pi_3 P \to \mathbb{Z}\pi_1 P \quad \text{ and } \quad \mu_3 : \pi_3 P \to \mathbb{Z}\pi_1 P / \langle g + g^{-1}, 1 \rangle.$$

The intersection invariant $\lambda_3$ can be computed geometrically by representing the two homotopy classes by transverse based maps $S^3 \to P$ and counting their intersection points with signs and group elements. Similarly, for the self-intersection invariant $\mu_3$ one represents the homotopy class by a generic map $a : S^3 \looparrowright P$ and counts self-intersections, again with signs and group elements:

$$\mu_3(a) := \sum_p \epsilon_p \cdot g_p$$

using a whisker to $a$ from the basepoint of $P$. We note that in this dimension, switching the choice of sheets at a double point $p$ changes $g_p \in \pi_1 P$ to $g_p^{-1}$ (as in dimension 4) but the signs change from $\epsilon_p$ to $-\epsilon_p$, explaining the relation $g + g^{-1} = 0$ in the range of $\mu_3$ (as a opposed to $g - g^{-1} = 0$ in the range of $\mu_2$ in dimension 4). The relation $1 = 0$ is important to make $\mu_3(a)$ only depend on the homotopy class of $a$ since a cusp homotopy introduces a double point with arbitrary sign and trivial group element (as in dimension 4). Changing the whisker for $a$ changes $\mu_3(a)$ by a conjugation with the corresponding group element. The homotopy invariance of $\mu_3$ follows from the fact that a generic homotopy is isotopic to a sequence of cusps, finger moves and Whitney moves, none of which changes the invariant.

Using the involution $\bar{g} := g^{-1}$ on $\mathbb{Z}\pi_1 P$, the "quadratic form" $(\lambda_3, \mu_3)$ satisfies the formulas

$$(**) \qquad \mu_3(a+b) = \mu_3(a) + \mu_3(b) + [\lambda_3(a,b)] \quad \text{and} \quad \lambda_3(a,a) = \mu_3(a) - \overline{\mu_3(a)}$$

where the second formula has no content for the coefficient at the trivial element in $\pi_1 P$: Since $\lambda_3$ is skew-hermitian, it vanishes on the left hand side, whereas it's not even defined on the right hand side.

The case $N = M \times \mathbb{R}$ of the following lemma describes the homomorphism used in Theorem 1.1 and will be used in the definition of the Freedman–Quinn invariant given in section 4.B. Recall that $T_N$ denotes the 2-torsion in $\pi_1 N$.

**Lemma 4.1.** *If $P^6 = N^5 \times I$, then $\mu_3 : \pi_3 N \to \mathbb{F}_2 T_N \leq \mathbb{Z}\pi_1 N/\langle g + g^{-1}, 1\rangle$ is a homomorphism.*

*Proof.* First note that the intersection pairing $\lambda_3$ vanishes identically, since one can represent $a, b \in \pi_3(N \times I)$ disjointly (and hence transversely without intersections) in $N \times 0$ respectively $N \times 1$. So from the second formula in $(**)$ above, together with the observation that $\mathbb{F}_2 T_N \leq \mathbb{Z}\pi_1 N/\langle g + g^{-1}, 1\rangle$ is the subgroup generated by $\{\zeta \in \mathbb{Z}\pi_1 N \mid \zeta = \bar{\zeta} \neq 1\}$, we see that $\mu_3(a)$ lies exactly in $\mathbb{F}_2 T_N$. And from the first formula in $(**)$ it follows that $\mu_3 : \pi_3 N \to \mathbb{F}_2 T_N$ is a homomorphism. $\square$

The next lemma will be used in the proof of Corollary 1.3 given in section 6.

**Lemma 4.2.** *$\mu_3$ factors through the Hurewicz homomorphism $\pi_3 P \twoheadrightarrow H_3 \widetilde{P}$.*

*Proof.* We will use Whitehead's exact sequence $\Gamma(\pi_2 P) \to \pi_3 P \twoheadrightarrow H_3 \widetilde{P}$ from [13], where the first map is induced by the quadratic map $\eta : \pi_2 P \to \pi_3 P$ which is pre-composition by the Hopf map $h : S^3 \to S^2$. We need to show that

$$\mu_3(a_3 + \eta(a_2)) = \mu_3(a_3) \ \forall \ a_i \in \pi_i P,$$

By the quadratic property of $\mu_3$ given by the first formula in $(**)$, we get

$$\mu_3(a_3 + \eta(a_2)) = \mu_3(a_3) + \mu_3(\eta(a_2)) + \lambda_3(a_3, \eta(a_2))$$

and so we want to show that the last two terms on the right vanish. Representing $a_2$ by an embedding $b_2 : S^2 \hookrightarrow P^6$, we see that $\eta(a_2) = b_2 \circ h$ is supported in the image of $b_2$. As a consequence of working in a 6–manifold, we can find a representative of $a_3$ in the complement of this 2–manifold and hence their intersection invariant $\lambda_3$ vanishes. Similarly, there is a generic representative of $\eta(a_2)$ which has support in the normal bundle of $b_2$, a simply-connected 6–manifold. Therefore, $\mu_3(\eta(a_2)) = 0$ since the trivial group element is divided out in the range of $\mu_3$. $\square$

**Remark 4.3.** Even though we obtain a map $\mu_3 : H_3\widetilde{P} \to \mathbb{Z}\pi_1 P/\langle g+g^{-1}, 1\rangle$, it is not clear to us whether $\mu_3$ can be computed in a "homological way", i.e. without representing homology classes by generic maps and counting double points. This can be done for $\lambda_3$ but the second formula in $(**)$ shows that $\lambda_3(a,a)$ does *not* determine $\mu_3(a)$ at group elements of order 2.

4.B. **The self-intersection invariant for homotopies of 2–spheres in 5–manifolds.** The above description of $\mu_3$ can also be applied to define self-intersection invariants of properly immersed simply-connected 3–manifolds in a 6–manifold. In this setting $\mu_3$ is computed just as above, by summing signed double point group elements, and is invariant under homotopies that restrict to isotopies on the boundary.

Now fix a smooth oriented 5–manifold $N$. For any homotopy $H : S^2 \times I \to N^5$ between embedded spheres in $N$ we define the self-intersection invariant of $H$

$$\mu_3(H) \in \mathbb{Z}\pi_1 N/\langle g + g^{-1}, 1\rangle$$

to be the self-intersection invariant $\mu_3$ of a generic track $S^2 \times I \looparrowright N^5 \times I$ for $H$ (with fixed boundary and based at the sphere $H_0$). The invariant $\mu_3(H)$ is independent of the choice of generic track since any two choices of perturbations to make $S^2 \times I \looparrowright N^5 \times I$ generic differ at most by a homotopy rel boundary.

Definition 4.5 of the Freedman–Quinn invariant below involves the case where $N^5 = M^4 \times \mathbb{R}$ and $H_0, H_1$ are embeddings $S^2 \hookrightarrow M \times 0$. In this case one has that $\mu_3(H) \in \mathbb{F}_2 T_M$, as in Lemma 4.1. The next lemma characterizes the dependence of $\mu_3(H)$ on the choice of homotopy $H$ only in this case, even though there is a more general formula explained in [8]:

**Lemma 4.4.** *If $J : S^2 \times I \looparrowright M \times \mathbb{R} \times I$ is a generic track of a based self-homotopy of $R : S^2 \hookrightarrow M \times 0$, then $\mu_3(J) \in \mathbb{F}_2 T_M$ lies in the image of the homomorphism $\mu_3 : \pi_3 M \to \mathbb{F}_2 T_M$.*

It follows that for any two based homotopies $H, H' : S^2 \times I \to M^4 \times \mathbb{R}$ between embedded spheres $H_0 = H_0'$ and $H_1 = H_1'$ in $M \times 0$, the difference $\mu_3(H) - \mu_3(H') \in \mathbb{F}_2 T_M$ lies in the image of $\mu_3 : \pi_3 M \to \mathbb{F}_2 T_M$, since stacking the two homotopies gives a based self-homotopy $J = H \cup -H'$ such that $\mu_3(J) = \mu_3(H) - \mu_3(H')$.

*Proof.* By assumption, $J$ agrees with the track $R \times I$ of the product self-homotopy on the 2-skeleton $S^2 \times \{0,1\} \cup z_0 \times I$ of $S^2 \times I$. So they only differ on the 3-cell where $R \times I$ is represented by $R(D^2) \times I$ (here $D^2$ is the complement in $S^2$ of a small disk around $z_0$) and $J$ is represented by a generic 3–ball $B : D^3 \looparrowright (M \times \mathbb{R} \times I) \smallsetminus \nu(z \times I)$. Here $z$ denotes the image of the basepoint $z_0 \in S^2$, and by construction the boundaries of these 3–balls are parallel copies of an embedded 2–sphere in the boundary of a small neighborhood of $R \times \{0,1\} \cup (z \times I)$. Gluing $B$ and $R(D^2) \times I$ together along a small cylinder $S^2 \times I$ between their boundaries yields a map of a 3–sphere $b := B \cup R(D^2) \times I : S^3 \to M \times \mathbb{R} \times I$. To prove the lemma we will show that $\mu_3(J) = \mu_3(b) \in \mu_3(\pi_3(M))$.

First note that all contributions to $\mu_3(J)$ come from double point loops in $B$. There are two types of self-intersections that contribute to $\mu_3(b)$, namely the self-intersections of the immersed 3–ball $B$ and the intersections between $B$ and the embedded 3–ball $R(D^2) \times I$. Observe that $B \pitchfork R(D^2) \times I = J \pitchfork R \times I$, with the corresponding loops based at $z \in R \times 0$ determining the same group elements contributing to both of $\mu_3(b)$ and $\lambda_3(J, R \times I)$.

Now note that $\lambda_3(J, R \times I) = 0$, since $R \times I \subset M \times 0 \times I$ can be made disjoint from a homotopic (rel boundary) copy of $J$ in $M \times 1 \times I$. So $B \pitchfork R(D^2) \times I$ contributes trivially to $\mu_3(b)$, and it follows that $\mu_3(b) = \mu_3(J)$ since both are determined by double point loops in $B$. $\square$

**Definition 4.5.** Given embeddings $R, R' : S^2 \hookrightarrow M^4$ which are based homotopic, their Freedman–Quinn invariant is given by:

$$\mathrm{fq}(R, R') := [\mu_3(H)] \in \mathbb{F}_2 T_M / \mu_3(\pi_3 M)$$

for any choice of based homotopy $H$ from $R \times 0$ to $R' \times 0$ in $M \times \mathbb{R}$.

Recall from the beginning of the proof of Lemma 2.1 that a common dual for $R$ and $R'$ forces any given homotopy in $M$ to be based and hence $\mathrm{fq}(R, R')$ is defined for any pair $R, R' \in \mathcal{R}^G_{[f]}$. This definition of $\mathrm{fq}(R, R')$ is independent of the choice of $H$ by Lemma 4.4.

4.C. **Computing the Freedman–Quinn invariant.** We show how to compute $\mathrm{fq}(R, R')$ as a "difference of sheet choices" for embedded 2-spheres $R \times 0$ and $R' \times 0$ in $M \times \mathbb{R}$.

Consider a homotopy $H$ given by finger moves on $R$ leading to a middle level $f : S^2 \looparrowright M$, followed by Whitney moves on $f$ leading to $R'$. The collection of Whitney disks $\mathcal{W}$ on $f$, inverse to the finger moves, gives $f_{\mathcal{W}} = R$ and determines a choice of sheets $\mathsf{x} = (x_1, \ldots, x_{2n})$, and the collection of Whitney disks $\mathcal{W}'$ such that $f_{\mathcal{W}'} = R'$ determines a choice of sheets $\mathsf{x}' = (x'_1, \ldots, x'_{2n})$.

We will describe an isotopy in $M \times \mathbb{R}$ from $R \times 0$ to $f \times b$, where $b : S^2 \to \mathbb{R}$ will be a sum of bump functions that "resolves" the double points in $f$. For simplicity of notation, we'll assume that $f$ is the result of just a single finger move, with $\mathsf{x} = (x_1, x_2)$.

First define for each $x \in S^2$ a smooth family of non-negative bump functions $b^x_s : S^2 \to \mathbb{R}$ which are supported in a small neighborhood of $x$ and have maximum $b^x_s(x) = s$. There is a homotopy $R_s$, $s \in [0, 1]$, describing how the finger grows from $R$ to the self-tangency which introduces an identification of $x, y \in S^2$, where $y$ gives the "finger tip" $R_s(y)$ while $R_s(x)$ is fixed for all $s$. It gives an isotopy $R_s \times b^x_s$ from $R \times 0$ to $R_1 \times b^x_1$, with the self-tangency avoided by the bump $b^x_1$ having lifted the image of the $x$-sheet above what was the tangency point (see Figure 20 left).

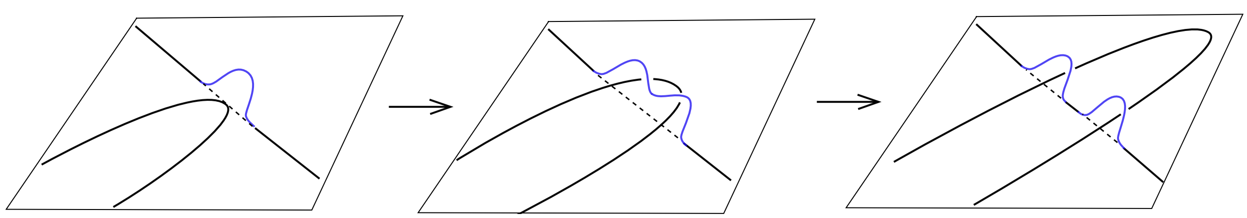

FIGURE 20. A single bump splitting into two, along a finger move.

We extend this to an isotopy in $M \times \mathbb{R}$ from $R \times 0$ to an embedding $f \times b$: As $R_s$ continues to move towards $f$, the self-tangency splits into two transverse intersection points, and we arrange the single bump $b^x_1$ to split into a sum of two bumps which finally arrives at $b := b^{x_1}_1 + b^{x_2}_1$ when the finger move is done, see Figure 20.

Note that in this convention, the chosen sheets $x_i \in S^2$ represent "over-crossings" of the embedding $f \times b$. The isotopy class of this embedding does not depend on the particulars of $b$ but only on the choice of sheets $\mathsf{x}$. In the general case of $n$ finger moves such a $b$ can be defined simultaneously to get a corresponding isotopy.

Turning the homotopy $H$ upside down, we can also consider finger moves leading from $R'$ to $f$ which are inverse to the Whitney moves along Whitney disks in $\mathcal{W}'$. Apply the same procedure using the choice of sheets $\mathsf{x}' = (x'_1, \dots, x'_{2n})$ to get an isotopy in $M \times \mathbb{R}$ from $R' \times 0$ to $f \times b'$. If $x_i = x'_i$ we have $b = b'$ near $x_i$, so these two isotopies can be glued together in that neighborhood.

If $x_i \neq x'_i$ there is a local homotopy $H_i(s) := f \times (b^{x'_i}_{1-s} + b^{x_i}_s)$ that moves $f \times b'$ to locally coincide with $f \times b$ by a "crossing change" (see Figure 21). $H_i$ has a single double point where it identifies $(x_i, 1/2)$ with $(x'_i, 1/2)$. The associated group element is $g(x_i) \in \pi_1 M$ associated to the sheet choice $x_i$ of the double point $f(x_i)$.

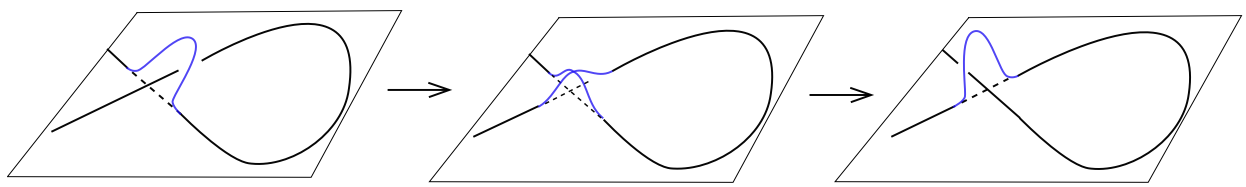

FIGURE 21. Two bumps crossing in a single point during a local homotopy $H_i$.

Assembling such local homotopies $H_i$ around all $x_i \neq x'_i$, and then composing with the above isotopies from $R \times 0$ to $f \times b$ and from $f \times b'$ to $R' \times 0$, yields a based homotopy $H_{\mathcal{W},\mathcal{W}'}$. Its isotopy class rel boundary only depends on the sheet choices $\mathsf{x}, \mathsf{x}'$ and not on the particulars of the bump functions in the construction.

**Lemma 4.6.** $\mu_3(H_{\mathcal{W},\mathcal{W}'}) = \sum_i g(x_i) \in \mathbb{F}_2 T_M$, *where the sum is over those double points* $p_i$ *of* $f$ *for which* $x_i \neq x'_i$. *This sum is therefore a representative for* $\operatorname{fq}(R, R') \in \mathbb{F}_2 T_M / \mu_3(\pi_3(M))$.

Recall from Lemma 4.1 and section 4.B that the target of $\mu_3(H_{\mathsf{x},\mathsf{x}'})$ is indeed the subgroup $\mathbb{F}_2 T_M$ of $\mathbb{Z}\pi_1 N / \langle g + g^{-1}, 1 \rangle$, i.e. any $g(x_i)$ with $g(x_i)^2 \neq 1$ must contribute trivially (and we don't have to worry about signs).

4.D. **Singular circles: The origin of the** fq **invariant.** The fq-invariant originally appeared in the more general setting of [2, Chap.10.9] as the obstruction to eliminating circles of intersections between the cores of 3-handles in a 5-manifold. For the interested reader we briefly explain the connection with singular circles in our setting. The results of this section will not be used in our paper.

The singular set of a generic track $S^2 \times I \looparrowright M \times I$ of a regular homotopy from $R$ to $R'$ consists of circles which are double-covered by circles in $S^2 \times I$. The group element associated to a singular circle is determined by a double point loop in the image of $S^2 \times I$ that changes sheets exactly at one point on the singular circle, with a choice of first sheet orienting the loop. The group element $g(\gamma)$ associated to a circle $\gamma$ with connected double cover satisfies $g(\gamma)^2 = 1$ since $\gamma$ itself represents $g(\gamma)$ and the double cover bounds a disk in the domain. The singular arcs that appear in [2, Chap.10.9] and start/end at cusps, do not occur in our setting since we work with a regular homotopy.

**Lemma 4.7.** $\operatorname{fq}(R, R') = [\sum_\gamma g(\gamma)] \in \mathbb{F}_2 T_M / \mu_3(\pi_3(M))$, *where the sum is over all* $\gamma$ *that have connected double covers in* $S^2 \times I$.

*Sketch of Proof:* The idea is to resolve the singular circles of a track $H : S^2 \times I \looparrowright M \times I$ to (at worst) self-intersection points of $S^2 \times I \looparrowright M \times \mathbb{R} \times I$, and compute $\mu_3$. Using the extra $\mathbb{R}$-factor, the singular circles with disconnected covers can be eliminated by perturbing one sheet into the $\mathbb{R}$-direction. By perturbing the sheets that intersect in a circle $\gamma$ with connected double cover partially into the positive $\mathbb{R}$-direction and partially into the negative $\mathbb{R}$-direction, $\gamma$ can be eliminated except for a single transverse self-intersection with group element $g(\gamma)$. $\square$

It is interesting to note that these singular circles in $M \times I$ project to the middle level $f : S^2 \times 1/2 \looparrowright M \times 1/2$ as follows: They map to the union of the boundary arcs of Whitney disks $W_i$ (inverse to finger moves on $R$) and the boundary arcs of Whitney disks $W_i'$ (guiding Whitney moves towards $R'$). These arcs meet at the self-intersections of $f$, so the union $\cup_i \partial W_i \cup \partial W_i'$ is a map of circles into $f(S^2)$. The number of circles will not in general be the number of self-intersection pairs, because the $W_i$ and $W_i'$ may induce different pairings.

To see that these Whitney disk boundaries are projections of the singular circles to the middle level $f$, consider first the track of the $i$th finger move: As the finger first touches the sheets and then pushes through, a single tangential self-intersection is created which then splits into two self-intersections that move apart until coming to rest at the end of the finger's motion. So in each sheet the motion of a single point splitting into two traces out one arc in the boundary of the Whitney disk $W_i$ (inverse to the finger move). In the domain $S^2 \times I$ of the homotopy we see neighborhoods of two minima of singular circles, see Figure 22. Turning the homotopy upside down, the same observations explain neighborhoods of the maxima.

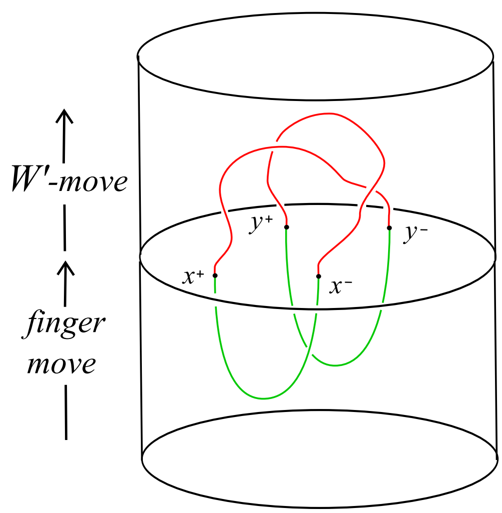


FIGURE 22. Singular circles in $S^2 \times I$: A connected double cover.

Singular circles with connected double covers arise when there are differences in the sheet choices determined by the $W_i$ and $W_i'$ as shown in Figure 22. This is consistent with our two computations of the Freedman–Quinn invariant in Lemmas 4.6 and 4.7: Each singular circle with double point loop $g$ corresponds to $n$ finger moves along the same $g$ and $n$ Whitney moves resolving the resulting double points. The number $n$ is the number of minima (and maxima) of the projection $M \times I \to I$ when restricted to the singular circle. The double cover is connected if and only if $g^2 = 1$ and there is an odd number of sheet changes from the sheet choice determined by the finger moves to the sheet choice of the Whitney moves.

## 5. Proof of Theorem 1.1

The last sentence of Theorem 1.1 follows from the fact that all our constructions, including throughout this section, are supported away from $G$. For the main part of Theorem 1.1, we will proceed with the following steps:

A. Define the geometric action of $\mathbb{F}_2 T_M$ on $\mathcal{R}^G_{[f]}$ and show that

$$\mathrm{fq}(t \cdot R, R) = [t] \in \mathbb{F}_2 T_M / \mu_3(\pi_3 M) \quad \forall R \in \mathcal{R}^G_{[f]}, t \in \mathbb{F}_2 T_M.$$

B. Show that the stabilizers are $\mu_3(\pi_3 M)$.
C. Prove that $R'$ is isotopic to $\mathrm{fq}(R, R') \cdot R$ for all $R, R' \in \mathcal{R}^G_{[f]}$.

The last item implies the transitivity of the action, so these steps complete the proof of Theorem 1.1: For a fixed $R \in \mathcal{R}^G_{[f]}$ the Freedman–Quinn invariant $\mathrm{fq}(R, \cdot) \in \mathbb{F}_2 T_M / \mu_3(\pi_3 M)$ inverts the $\mathbb{F}_2 T_M$-action. $\square$

5.A. **The geometric action on $\mathcal{R}^G_{[f]}$.** An outline of this construction was given in section 1.C. Given $t = t_1 + \cdots + t_n \in \mathbb{F}_2 T_M$ and $R \in \mathcal{R}^G_{[f]}$, we first do $n$ finger moves on $R$, along arcs starting and ending near the base-point in $R$, representing $t_i \in T_M$. The isotopy class of the resulting generic map $f^t : S^2 \looparrowright M$ only depends on $R$ and $t$ because $\pi_1(M \smallsetminus R) \cong \pi_1 M$ and homotopy implies isotopy for arcs in 4–manifolds.

The second step in the definition of our action is to do Whitney moves on $f^t$ along a collection $\mathcal{W}^t$ of $n$ Whitney disks to arrive at an embedding denoted by $t \cdot R$, where $\mathcal{W}^t$ satisfies the following sheet choice condition: Let $\mathsf{x} = (x_1^+, x_1^-, \ldots, x_n^+, x_n^-)$ be a sheet choice such that the collection $\mathcal{W}$ of Whitney disks $W_i$ which are inverse to the finger moves is $\mathsf{x}$-compatible and each $W_i$ pairs $f(x_i^\pm)$, i.e. $\mathcal{W}$ is also compatible with the pairing choice $\mathsf{x}^\pm = (x_1^\pm, \ldots, x_n^\pm)$. Then we take $\mathcal{W}^t$ to be any choice of Whitney disks that is compatible with the sheet choice $\mathsf{x}^t := (x_1^+, y_1^-, \ldots, x_n^+, y_n^-)$ which has the sheets of $f^t$ switched at each negative self-intersection $f^t(x_i^-) = f^t(y_i^-)$. Such an $\mathsf{x}^t$-compatible $\mathcal{W}^t$ exists by Lemma 3.1, and by Corollary 3.6 the isotopy class of $f^t_{\mathcal{W}^t}$ is determined by $\mathsf{x}^t$, so $t \cdot R := f^t_{\mathcal{W}^t} \in \mathcal{R}^G_{[f]}$ is well defined. Lemma 4.6 implies by construction:

**Lemma 5.1.** $\mathrm{fq}(t \cdot R, R) = [t]$ *for all* $R \in \mathcal{R}^G_{[f]}$ *and* $t = t_1 + \cdots + t_n \in \mathbb{F}_2 T_M$.

By Corollary 3.10, sheet choices $\mathsf{x}$ don't effect the isotopy class of $f_{\mathsf{x}}$ at double points whose group element is not 2-torsion. This implies that $t \cdot R$ is unchanged if we perform more finger moves on $R$ along non-2-torsion (and then appropriate Whitney moves to arrive at an embedding). In Lemma 3.7 we showed that making double sheet changes doesn't change the isotopy class of $f_{\mathsf{x}}$, so only the mod 2 number of finger moves along 2-torsion matters for the isotopy class of $t \cdot R$:

**Lemma 5.2.** *For* $R \in \mathcal{R}^G_{[f]}$ *and* $t = t_1 + \cdots + t_n \in \mathbb{F}_2 T_M$, $t \cdot R = R' \in \mathcal{R}^G_{[f]}$ *for any* $R'$ *that is obtained from* $R$ *by a sequence of finger moves and Whitney moves as long as* $\mu_3(H_{\mathcal{W},\mathcal{W}'}) = t$.

Recall that by Lemma 4.6 $\mu_3(H_{\mathcal{W},\mathcal{W}'}) = \sum_{x_i \neq x_i'} g(x_i)$ only depends on the middle level of the homotopy and the two sheet choices $\mathsf{x}$ and $\mathsf{x}'$ (and only at double points whose group elements are 2-torsion and which are counted mod 2).

5.B. **The stabilizer equals $\mu_3(\pi_3 M)$.**

**Lemma 5.3.** *If $t \cdot R$ is isotopic to $R$, then $t \in \mu_3(\pi_3 M)$, i.e. the stabilizer of $R \in \mathcal{R}^G_{[f]}$ is contained in $\mu_3(\pi_3 M)$.*

*Proof.* The union of a based homotopy $H^t$ from $R$ to $t \cdot R$ with $\mu_3(H^t) = t$ and a based isotopy $H^0$ from $t \cdot R$ to $R$ forms a based self-homotopy $J := H^t \cup H^0$ of $R$. So by Lemma 4.4, we have $t = \mu_3(H^t) = \mu_3(J) \in \mu_3(\pi_3 M)$. □

**Lemma 5.4.** *If $t \in \mu_3(\pi_3 M)$ then $t \cdot R$ is isotopic to $R$, i.e. $\mu_3(\pi_3 M)$ is contained in the stabilizer of any $R \in \mathcal{R}^G_{[f]}$.*

*Proof.* We first use that a closed tubular neighborhood $\nu(R \cup G)$ has boundary $S^3$ and is homotopy equivalent to $S^2 \vee S^2$ (in fact, capping it off with $B^4$ leads to a sphere-bundle over $S^2$ with Euler number $R \cdot R$). If $M_0 \subset M$ is the closure of the complement of $\nu(R \cup G)$ then the corresponding Mayer-Vietoris sequence (for universal covering spaces) reads as follows:

$$H_3(\nu(R \cup G); \mathbb{Z}\pi_1 M) \oplus H_3(\widetilde{M_0}) \longrightarrow H_3(\widetilde{M}) \longrightarrow H_2(S^3; \mathbb{Z}\pi_1 M)$$

Since the first and last terms are 0, we see that the inclusion induces an epimorphism $H_3(\widetilde{M_0}) \twoheadrightarrow H_3(\widetilde{M})$. By the surjectivity of Hurewicz maps, this implies that we may assume that $t = \mu_3(a)$ for some $a \in \pi_3 M_0$.

Now represent $a$ by a based generic regular homotopy $F_s : S^2 \times I \to M_0$ from the trivial sphere $F_0 = F_1$ in $M_0$ to itself. By construction, $F_s$ lies in the complement of $R$ at each $s$-level, so we can take a smooth family of ambient connected sums of $F_s$ with $R \times s$ to get a homotopy $H : S^2 \times I \to M$ from $R$ to itself with $\mu_3(H) = t$. By Lemma 5.2, this shows that $F_1 \# R$ is an admissible representative of our action $t \cdot R$ and therefore, $t \cdot R$ is isotopic to $R$. □

5.C. **The action is transitive.** This follows directly from:

**Lemma 5.5.** *For any $R, R' \in \mathcal{R}^G_{[f]}$, we have* $\operatorname{fq}(R, R') \cdot R = R'$.

*Proof.* This is a simple consequence of Lemmas 4.6 and 5.2. □

## 6. Proofs of Corollaries 1.3 and 1.7

We first note that in the example given below Corollary 1.3, $\mu_3(\pi_3 M) = 0$ since $M$ (and hence its universal covering $\widetilde{M}$) has no 3-handles, and $\mu_3$ factors through the Hurewicz homomorphism $\pi_3(M) \twoheadrightarrow H_3(\widetilde{M}) = 0$ by Lemma 4.2. So $|\mathbb{F}_2 T_M / \mu_3(\pi_3 M)| = |\mathbb{F}_2 T_M| = \infty$.

The pseudo-isotopy statement of Corollary 1.3 follows from Lemma 6.1 below because a diffeomorphism $\varphi : M \times I \cong M \times I$ with $\varphi_0 = \mathrm{id}$ (the pseudo-isotopy condition) and $\varphi_1(R) = R'$ leads to the concordance $\varphi \circ (R \times \mathrm{id}) : S^2 \times I \hookrightarrow M \times I$ from $R$ to $R'$. This contradicts $\operatorname{fq}(R, R') \neq 0$ by Corollary 7.1. □

**Lemma 6.1.** *Let $G : S^2 \hookrightarrow M$ be framed and fix $n \in \mathbb{Z}$. Then the diffeomorphism group of $M$ acts transitively on embedded spheres $R : S^2 \hookrightarrow M$ with $G$ as a geometric dual and normal Euler number $e(\nu R) = n$.*

*Proof.* Given $G, R$ as above, consider a closed regular neighborhood $\nu(R \cup G) \subset M$. It is diffeomorphic to the 4-manifold $M_n$ with one 0-handle and two 2-handles attached to the

Hopf link, one 0-framed and the other $n$-framed. In particular, the boundary $\partial M_n$ is a 3-sphere which leads to a decomposition

$$M \cong M_n \cup_{S^3} M_R,$$

where $M_R$ is the closure of the complement of $M_n$ in $M$. Note that $G : S^2 \hookrightarrow M_n \subset M$ is the union of the (core of the) 0-framed 2-handle and a disk bounding the 0-framed component of the Hopf link. As a consequence, surgery on $G$ in $M_n$ leads to the 4-manifold where that 0-framed 2-handle is replaced by a 1-handle. This 1-handle then cancels the $n$-framed 2-handle, showing that surgery on $G$ leads from $M_n$ to $D^4$. It follows that surgery on $G$ also leads from $M$ to $M_R \cup_{S^3} D^4$.

Repeating the same constructions for $R'$ in place of $R$, we get a second decomposition

$$M \cong M_n \cup_{S^3} M_{R'},$$

where $M_{R'} \cup_{S^3} D^4$ is diffeomorphic to surgery on $G$ in $M$. But $G$ is a *common* dual, so we get an orientation preserving diffeomorphism $M_R \cong M_{R'}$. Since orientation preserving diffeomorphisms of $S^3$ are isotopic to the identity, we can extend this to a self-diffeomorphism of $M$ which carries $R$ to $R'$ and fixes $G$: This just requires to line up the 2-handles of $M_n$ in the obvious way. $\square$

The proof of Corollary 1.7 comes about as follows: For $t \in \mathbb{F}_2 T_M$, the relative unknotting number satisfies $\mathrm{u}(t \cdot R, R) \leq |t|$ because $t \cdot R$ is constructed from $R$ by using $|t|$ finger moves. Moreover, any $R' \in \mathcal{R}^G_{[f]}$ is isotopic to some $t \cdot R$, so it suffices to understand those particular numbers. If $[t] = [s]$ then $t \cdot R = [t] \cdot R$ is isotopic to $s \cdot R$, so $\mathrm{u}(t \cdot R, R) \leq |s|$ holds as well.

If $u := \mathrm{u}(t \cdot R, R)$ then there are $u$ finger moves and then $u$ Whitney moves that lead from $R$ to $t \cdot R$. By general position, we may assume that the finger moves are disjoint from $G$ and run along group elements $g_i \in \pi_1 M, i = 1, \ldots, u$. By Lemma 2.4 we find Whitney disks with the same sheet choices in the complement of $G$, and by Lemma 5.2 they also lead to $t \cdot R$. This implies that $u$ is at least as large as the number of 2-torsion $s_j$ among the $g_i$ which by itself equals $|s|$ for $s := \sum_j s_j$. So we get $u \geq |s|$ and together $u = |[t]|$ as claimed. $\square$

## 7. Ambient Morse theory and the $\pi_1$-negligible embedding Theorem

A third proof of Gabai's LBT arises from ambient Morse theory and the uniqueness part of the $\pi_1$-negligible embedding theorem [2, 11, Thm.10.5A(2)]. We state it in the orientable setting and recall that an embedding $h : V \hookrightarrow W$ is $\pi_1$-*negligible* if the inclusion induces an isomorphism $\pi_1(W \smallsetminus h(V)) \cong \pi_1 W$.

**Theorem 10.5(2).** *Let $(V; \partial_0 V, \partial_1 V)$ be a compact 4–manifold triad so that $\pi_1(V, \partial_0 V) = \{1\} = \pi_1(V, \partial_1 V)$ (all basepoints), each component has nonempty intersection with $\partial_1 V$, and components disjoint from $\partial_0 V$ are 1-connected.*

*Suppose $W$ is an oriented 4–manifold, $h, h' : (V, \partial_0 V), \hookrightarrow (W, \partial W)$ are $\pi_1$-negligible embeddings, and $H$ is a homotopy rel $\partial_0 V$. Then there is an obstruction $\mathrm{fq}(H) \in H^2(V, \partial_0 V; \mathbb{F}_2 T_W)$ which vanishes if and only if $H$ is homologous (with $\mathbb{Z}[\pi_1 W]$-coefficients) to a $\pi_1$-negligible concordance $V \times I \hookrightarrow W \times I$ from $h$ to $h'$.* $\square$

This is the statement given in [11, p.2], where the notation for the obstruction is introduced and the dependence on $H$ is pointed out. Stong then continues to correct this statement by showing that in general, there is a secondary obstruction, the *Kervaire-Milnor invariant*, to finding a concordance. It is only relevant if $h$ is s-characteristic, which we'll show not to be the case in our application (because our dual $G$ is framed). Stong also observes on the bottom of page 2 that fq can be strengthened to be independent of $H$ by taking $\mathrm{fq}(h, h')$ in the quotient of $H^2(V, \partial_0 V; \mathbb{F}_2 T_W)$ by the self-intersection invariant on $\pi_3 W$. Note that this is a 5-dimensional result so it holds in the smooth category.

We apply this theorem for $W$ defined to be the manifold $M$, with an open neighborhood of $G$ removed, and $V := D^2 \times D^2$ with $\partial_0 V = S^1 \times D^2$ and $\partial_1 V = D^2 \times S^1$. Then $R, R'$ can be turned into embeddings $h, h' : (V, \partial_0 V) \hookrightarrow (W, \partial W)$ by using the normal bundles of $R, R'$ and removing a neighborhood of their intersection point with $G$. Note that $R$ may have non-trivial normal bundle (necessarily isomorphic to that of $R'$) but after removing the neighborhood of $G$, it turns into a $D^2$-bundle over $D^2$ which must be trivial.

By Lemma 2.1, the resulting embeddings $h, h'$ are homotopic rel $\partial_0 V$ and the theorem applies. Note that $(V, \partial_0 V) \simeq (D^2, S^1)$ and hence the invariant $\mathrm{fq}(R, R') = \mathrm{fq}(h, h')$ lies in $\mathbb{F}_2 T_W / \mu_3(\pi_3 W)$. Note also that Seifert–van Kampen shows that in this case, every concordance is $\pi_1$-negligible (as long as it is on one boundary).

If $\mathrm{fq}(h, h') = 0$ then $h$ and $h'$ are concordant by the above theorem. We now reverse the above steps of thickening spheres and disks to 4–manifolds with boundary to arrive at a concordance $C : S^2 \times I \hookrightarrow M \times I$ between $R$ and $R'$ as in the corollary below. Note that Stong's additional Kervaire-Milnor invariant vanishes in our setting since $R$ is not s-characteristic: The dual sphere $G$ is framed, so that

$$R \cdot G \equiv 1 \neq 0 \equiv G \cdot G \mod 2.$$

**Corollary 7.1.** *Given embedded spheres $R, R' \in \mathcal{R}^G_{[f]}$ as in Theorem 1.1, the obstruction $\mathrm{fq}(R, R') \in \mathbb{F}_2 T_M / \mu_3(\pi_3 M)$ vanishes if and only if there is a concordance $C : S^2 \times I \hookrightarrow M \times I$ between $R$ and $R'$. Moreover, if $C$ exists one can arrange that it has $G$ as a geometric dual in every level $t \in I$: $C^{-1}(G \times \{t\}) = (z_0, t)$.* □

By the following result, which will be proven using ambient Morse theory and only basic lemmas from this paper, the Freedman–Quinn invariant completely detects isotopy in this setting:

**Theorem 7.2.** *Given a concordance $C : S^2 \times I \hookrightarrow M \times I$ between $R$ and $R'$ which has $G$ as a geometric dual in every level $t \in I$ as in the above corollary, it follows that $R$ and $R'$ are isotopic.*

*Proof.* We now show how to directly turn the concordance $C$ into an isotopy using the geometric duals. By general position, we may assume that the composition $p_2 \circ C : S^2 \times I \to I$ is a Morse function. If it has no critical points then $C$ is an isotopy, so we'll study the critical points by ambient Morse theory, compare [1, Sec.5]. In Lemma 8 of that paper, it was shown that by an ambient isotopy of $M \times I$ one can order the critical points according to their index.

Moreover, one can re-order critical points of the same index arbitrarily which can be seen as follows, say in the case of 1-handles: The core of a 1-handle (in the 3-manifold $S^2 \times I$) is an arc, whereas the cocore is a 2-disk. If we have two adjacent 1-handles just below, respectively above, a level $M = M \times t$ then we can push the cocore up and the core down into that "middle" level $M$. By general position, these 1- respectively 2-manifolds will not intersect in the ambient 4-manifold $M$ and hence we can push the upper 1-handle below the lower one.

As a consequence, we can assume that our Morse function on $S^2 \times I$ first has $n$ minima (0-handles) which are then abstractly cancelled by $n$ 1-handles: Each 0-handle must be abstractly cancelled eventually and we can slide those cancelling 1-handles below the other 1-handles. Looking at the top, $m$ maxima arise that are abstractly cancelled by $m$ 2-handles. The remaining 1- and 2-handles form a third cobordism which must be diffeomorphic to $S^2 \times I$ since gluing $S^2 \times I$ to its top and bottom gives the entire cobordism $S^2 \times I$.

More precisely, we can find two non-critical levels $t_1 < t_2$ in $(0, 1)$ such that $C^{-1}(M \times \{t_i\})$ are spheres which separate the domain $S^2 \times I$ of $C$ into three product cobordisms:

$$V_i := C^{-1}(M \times [t_i, t_{i+1}]),\ i = 0, 1, 2 \text{ and } t_0 := 0, t_3 := 1.$$

Here $V_i \cong S^2 \times I$ consists of the $i$- and $(i + 1)$-handles discussed above. Our proof will be completed by showing that each of the three restrictions of $C$ to $V_i$ can be turned into an isotopy, using the geometric dual $G$. $\square$

For $V_0$, the $t$-parameter gives a movie in $M$ that starts with $R$ and then shows $n$ trivial spheres $S_1, \dots, S_n$ being born in $M$, one for each 0-handle. Then $n$ tubes form, one for each 1-handle, that connect $R$ to each $S_i$, making the result a new sphere $R_w$ in $M$. Here $w$ is a collection of $n$ words in the free product $\pi_1 M * F_n \cong \pi_1(M \setminus \cup_i S_i)$, where $F_n$ is the free group generated by the meridians $m_i$ to $S_i$, and the words in $w$ measure how the core arcs of the 1-handles hit the cocore 3–balls of $S_i$ in $M$.

These cocores and cores originally lie in $M \times [0, t_1]$ but we pushed the cocores up and the cores down into a common middle level $M = M \times t_1/2$. By the above reordering argument, the collection $\mathcal{C}$ of cocores is embedded disjointly into $M$ and similarly, the collection $\mathcal{C}'$ of cores is also embedded disjointly. However, these 3– respectively 1–manifolds can intersect each other in the 4-dimensional middle level $M$, so the abstract handle cancellation can a priori not be done ambiently in $M$.

**Lemma 7.3.** *The sphere $R_w$ is isotopic to $R$ in any neighborhood of $R \cup \mathcal{C} \cup \mathcal{C}' \cup G$ in $M$.*

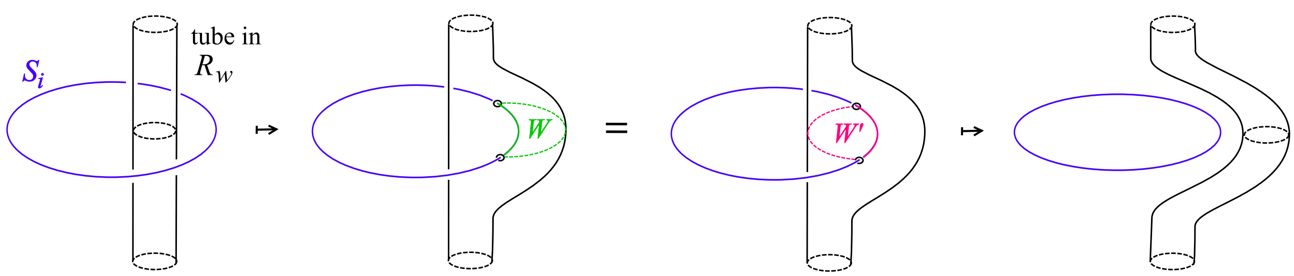


FIGURE 23. Pushing core arcs out of cocore 3-balls

*Proof.* Figure 23 shows how we can reduce the number of occurences of the meridian $m_i$ in $w$. This is a finger move and then a Whitney move on $R_w$, and as usual we see two Whitney disks, $W$ going back to $R_w$ by the inverse of the finger move and $W'$ going forward. These Whitney disks share a boundary arc $\beta$, and by Proposition 2.11 it follows that $R_w$ is isotopic to the result $R_{w'}$ of the Whitney move along $W'$, with $w'$ containing one letter $m_i$ less then $w$. Iterating this procedure we see that $R_w$ is isotopic to $R_{w_0}$ where $w_0 \in \pi_1 M$. This means that the 1-handles for $R_{w_0}$ do not intersect the cocore 3–balls for the 0-handles. These 3–balls then provide the final isotopy from $R_{w_0}$ to $R$. $\square$

Applying the same arguments of Lemma 7.3 to $V_2$ turned upside down shows that the restriction of $C$ to $V_2$ can be replaced by an isotopy. So it just remains to show that the restriction of $C$ to $V_1$ can be replaced by an isotopy.

The $t$-parameter movie for $V_1$ starts with the sphere $R$ at $t = t_1$ then $g$ tubes form, one for each remaining 1-handle in $V_1$. We then see a surface $F$ of genus $g$ in the middle level $M$ in which the collection $\mathcal{C}$ of cocores is also embedded. These are 2–disks, or better, a collection of $g$ caps attached to a half-basis of disjointly embedded simple closed curves in $F$. The movie continues with $g$ 2-handles being attached to $F$ whose cores form a second collection of caps $\mathcal{C}'$, again embedded disjointly into the middle level $M$.

**Lemma 7.4.** *The sphere $R$ is isotopic to $R'$ in any neighborhood of $F \cup \mathcal{C} \cup \mathcal{C}' \cup G$ in $M$.*

*Proof.* By construction, we have a genus $g$ surface $F \subset M$, together with a collection $\mathcal{C}$ of $g$ caps such that surgery leads to $R$, and another collection $\mathcal{C}'$ of $g$ caps for $F$ that surger it to $R'$. The caps in each collection are embedded in $M$, and disjoint from all other caps in the same collection, but caps of different collections may intersect on their boundary (in $F$) as well as in their interiors.

There are two handle-bodies $Y$ and $Y'$ formed from $F \times [-\epsilon, \epsilon]$ by (abstractly) attaching thickened caps from $\mathcal{C}$ to $F \times -\epsilon$, respectively $\mathcal{C}'$ to $F \times \epsilon$, and then filling the resulting boundary with two 3–balls. This is a Heegaard decomposition of $S^3$ to which we will next apply some classical 3–manifold results to simplify the intersection pattern in $F$ between the boundaries of the caps in $\mathcal{C}$ and those in $\mathcal{C}'$.

Waldhausen's uniqueness theorem for Heegaard decompositions of $S^3$ [12] gives a diffeomorphism of triples (isotopic to the identity – but we won't use this here)

$$(S^3; Y, Y') \cong (S^3; Y_0, Y_0')$$

where the subscript 0 refers to the standard Heegaard decomposition, stabilized to be of the same genus as $Y$. In the following, we'll need the usual notion of *minimal systems of disks*, which are disjointly embedded disks that cut a handlebody into a 3–ball. For $Y$, respectively $Y'$, such minimal systems are given by the caps in $\mathcal{C}$, respectively $\mathcal{C}'$. On the $(Y_0, Y_0')$-side these are standard disks in the sense that their boundaries meet $\delta_{ij}$ geometrically. By applying Waldhausen's diffeomorphism, we see that $Y$ and $Y'$ admit minimal systems of disks that also meet $\delta_{ij}$ geometrically on the boundary.

A result of Reidemeister [6] and Singer [10] from 1933 asserts that any two minimal systems of disks in a handlebody are slide equivalent. This implies that after finitely many handle slides among the abstract caps in $\mathcal{C}$ respectively $\mathcal{C}'$, we may assume that the collections of caps $\mathcal{C}$ and $\mathcal{C}'$ meet $\delta_{ij}$ on the boundary. These handle-slides can be achieved ambiently in $M$ and we'll assume from now on that this has been done. This has the consequence that the complement in $F$ of the boundaries of the caps in $\mathcal{C}$ and $\mathcal{C}'$ is connected. In particular, in the following arguments we may always find (disjoint) arcs from any point in this complement to the intersection point of $F$ and $G$.

If the interiors of all caps happen to be disjoint then Lemma 2.6 shows that the two surgeries $R$ and $R'$ are isotopic in $M$. We will complete our proof of Lemma 7.4 by showing the following general result. □

**Lemma 7.5.** *Let $F$ be a surface in a 4–manifold admitting two collections $\mathcal{C}, \mathcal{C}'$ of caps which are separately disjointly embedded, meet $\delta_{ij}$ on their boundary (in $F$), but can have* *interior intersections with each other. If $F$ has a geometric dual $G$ which is disjoint from $\mathcal{C}, \mathcal{C}'$ then there exists a collection $\mathcal{C}''$ with the same boundaries as $\mathcal{C}$, which has no interior intersections with $\mathcal{C}'$, and such that surgery on $\mathcal{C}''$ is isotopic to surgery on $\mathcal{C}$.*

Note that Lemma 2.6 then implies that surgery on $\mathcal{C}$ is also isotopic to surgery on $\mathcal{C}'$, which we wanted to prove.

*Proof.* Our construction will eliminate each intersection point $p \in c_i \pitchfork c'_j$ for $c_i \in \mathcal{C}$ and $c'_j \in \mathcal{C}'$ by tubing $c_i$ into a dual sphere $S_j$ to $c'_j$. This does not change $F_{\mathcal{C}'}$ since $\mathcal{C}'$ is fixed, and it will be checked that the tubing of the $c_i$ into the $S_j$ does not change $F_{\mathcal{C}}$ up to isotopy.

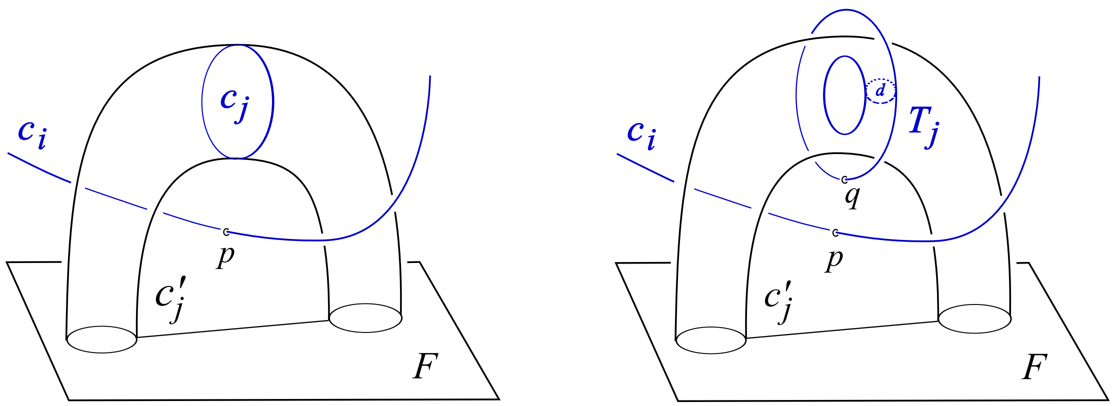


FIGURE 24. Left: An intersection $p \in c_i \pitchfork c'_j$. Right: A torus $T_j$ of normal circles over $\partial c_j$ with $T_j \cap c'_j = \{q\}$.

We first describe the easiest case where $\mathcal{C} \pitchfork \mathcal{C}'$ is a single interior intersection $p \in c_i \pitchfork c'_j$ for some $c_i \in \mathcal{C}$ and $c'_j \in \mathcal{C}'$ with $i \neq j$ (Figure 24, left). By assumption there exists a cap $c_j \in \mathcal{C}$ whose boundary $\partial c_j$ intersects $\partial c'_j$ in a single point. A torus $T_j$ of normal circles to $F$ over $\partial c_j$ intersects the interior of $c'_j$ in a single point $q$ (Figure 24, right). Let $d$ be a meridional disk to $F$ bounded by a circle in $T_j$, and denote by $d_G$ the result of tubing $d$ into $G$ to eliminate the intersection between $d$ and $\partial c_j$ (as in Figure 8 but here $\partial d \subset T_j$). Then surgering $T_j$ along $d_G$ yields a 0-framed embedded sphere $S_j$ with $q = S_j \cap c'_j$, such that $S_j$ is disjoint from all other caps in $\mathcal{C}'$, and $S_j$ is disjoint from all caps in $\mathcal{C}$ (Figure 25, left). So the intersection $p$ can be eliminated by tubing $c_i$ into $S_j$ along a path between $p$ and $q$ in $c_j$ (Figure 25, right).

At this point we have eliminated $p \in c_i \pitchfork c'_j$ by replacing $c_i$ with the connected sum $c''_i := c_i \# S_j$ of $c_i$ with $S_j$ to get a new collection of caps $\mathcal{C}''$ with the same boundaries as $\mathcal{C}$ but with interiors disjoint from $\mathcal{C}'$. We want to check that $F_{\mathcal{C}}$ is isotopic to $F_{\mathcal{C}''}$. Note that $T_j$ also admits a cap $\gamma_j$ formed from $c_j$ by deleting a small collar. (The boundary of $\gamma_j$ is visible in the right side of Figure 24 as the "inner longitude" of $T_j$.) This cap $\gamma_j$ is disjoint from $F$ and is dual to $d_G$, so it follows from the capped surface isotopy lemma (Lemma 2.6) that the sphere $S^\gamma_j$ formed by surgering $T_j$ along $\gamma_j$ is isotopic to $S_j$ in the complement of $F$. So it suffices to check that $F_{\mathcal{C}}$ is isotopic to $F_{\mathcal{C}^\gamma}$, where the collection of caps $\mathcal{C}^\gamma$ differs from the original $\mathcal{C}$ by replacing $c_i$ with $c_i \# S^\gamma_j$.

The sphere $S^\gamma_j$ is contained in the boundary of a tubular neighborhood $\nu_{c_j} \cong D^2 \times D^2$ of $c_j$, and $S^\gamma_j$ bounds an embedded 3–ball $B^\gamma_j \subset \nu_{c_j}$ which is the union of the solid torus $\partial c_j \times D^2$

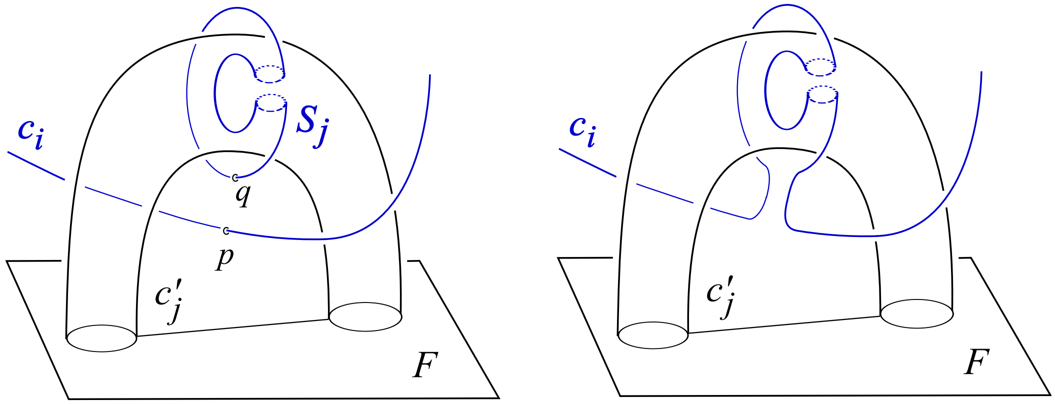


FIGURE 25. Left: The sphere $S_j$ with $S_j \cap F = \{q\}$. Right: The result of tubing $c_i$ into $S_j$ to eliminate $p$ and $q$.

with a 1-dimensional sub-bundle over the interior of $c_j$. Observe that the only intersections between $B_j^\gamma$ and $F$ are the circle $\partial c_j$.

Now surger $F$ along $\mathcal{C}^\gamma$ to get $F_{\mathcal{C}^\gamma}$. Since surgery has deleted a regular $\epsilon$-neighborhood of $\partial c_j$ from $F$, the 3–ball $B_j^\gamma$ is now disjoint from $F_{\mathcal{C}^\gamma}$. So there exists an isotopy from $F_{\mathcal{C}^\gamma}$ to $F_{\mathcal{C}}$ supported near $B_j^\gamma$ which isotopes the two parallel copies of $c_i \# S_j^\gamma$ in $F_{\mathcal{C}^\gamma}$ to the two parallel copies of $c_i$ in $F_{\mathcal{C}}$ by shrinking the parallels of $S_j^\gamma$ in $B_j^\gamma$.

The description of how this construction can be carried out in the general case to simultaneously eliminate any number of intersections $p \in c_i \pitchfork c'_j$ among all the $c_i \in \mathcal{C}$ and $c'_j \in \mathcal{C}'$ is straightforward: Consider some $c'_j$ which has multiple interior intersections with multiple $c_i$ (in the left of Figure 24 imagine more $p$-intersections). We will not introduce sub-index notation to enumerate the interior intersections in each $c'_j$, nor for the subsequent tori and spheres created for each intersection. Take a torus $T_j$ as in the right of Figure 24 around a parallel copy of $\partial c_j$ for each interior intersection. (Note that these parallels of $\partial c_j$ and their corresponding disjoint normal tori can be assumed to be supported arbitrarily close to $\partial c_j$, ie. in the part of $F$ that will be deleted by surgery – this observation is key to why the general case will present no new difficulties.) Just as above, these tori can be surgered to spheres $S_j$ disjoint from $F$ which are dual to $c'_j$ using caps $d_G$ on the $T_j$ in the complement of $F$ created by tubing meridional disks into $G$ along disjointly embedded arcs in $F$. These $S_j$ are all disjointly embedded by construction. Now all intersections between $c'_j$ and the $c_i$ can be eliminated by tubing the $c_i$ into the $S_j$ along disjointly embedded arcs in $c'_j$ between pairs of intersection points in $c_i \pitchfork c'_j$ and $S_j \cap c'_j$ (as in the right of Figure 25). Note that the case $i = j$ is allowed in this construction since the tori are supported near the parallel copies of $\partial c_j$ and the $S_j$ are disjoint from all $c_i$, so changing the interior of $c_j$ by tubing into an $S_j$ can be carried out just as for $c_i$ with $i \neq j$. Carrying out this construction for all $c'_j$ replaces $\mathcal{C}$ with $\mathcal{C}''$ such that $\mathcal{C}''$ and $\mathcal{C}'$ have disjoint interiors (with boundaries unchanged).

It remains to check that the argument from the easy case also applies to show that this construction which has changed the $c_i$ by multiple connected sums has not changed the result of surgery. As before, we can surger each of the $T_j$-tori along a cap $\gamma_j$ formed from a parallel of $c_j$ to get a sphere $S_j^\gamma$ which is isotopic in the complement of $F$ to the corresponding $S_j$. Here we are using parallels of the new $c''_j$ which may been tubed into some $S_k$'s, but the key properties of being framed, with interiors disjointly embedded in the complement of $F$ have been preserved. Since the $\gamma_j$-caps are dual to the $d_G$-caps, the $S_j^\gamma$-spheres are isotopic to the $S_j$-spheres in the complement of $F$, again by the capped surface isotopy lemma (Lemma 2.6). So again it suffices to check that $F_{\mathcal{C}}$ is isotopic to $F_{\mathcal{C}^\gamma}$ where the collection of caps $\mathcal{C}^\gamma$ differs from the original $\mathcal{C}$ by taking connected sums of the $c_i$ with multiple $c_i\#S_j^\gamma$.

Similarly as before, the $S_j^\gamma$ are contained in the boundaries of disjoint tubular $D^2 \times D^2$-neighborhoods of parallels of $c_j$, with each of these neighborhoods containing an embedded 3–ball $B_j^\gamma$ bounded by $S_j^\gamma$ such that $B_j^\gamma$ and $F$ only intersect in the corresponding parallel copy of $\partial c_j$. Surgering $F$ along $\mathcal{C}^\gamma$ to get $F_{\mathcal{C}^\gamma}$ deletes regular $\epsilon$-neighborhoods of all the $\partial c_j$ from $F$, and since we may assume that all the $T_j$-tori in the construction were supported near parallels of the $\partial c_j$ that lie inside these deleted $\epsilon$-neighborhoods, all the $B_j^\gamma$-balls are disjoint from $F_{\mathcal{C}^\gamma}$. So there exists an isotopy from $F_{\mathcal{C}^\gamma}$ to $F_{\mathcal{C}}$ supported near the $B_j^\gamma$ which isotopes the pairs of parallel copies of $c_i\#S_j^\gamma$ in $F_{\mathcal{C}^\gamma}$ to the pairs of parallel copies of $c_i$ in $F_{\mathcal{C}}$ by shrinking the parallels of $S_j^\gamma$ in $B_j^\gamma$. □

## References


[1] **A Bartels**, **P Teichner**, *All 2-dimensional links are null homotopic*, Geom. Topol., 3 (1999), 235–252.

[2] **M Freedman**, **F Quinn**, *The topology of* 4*-manifolds*, Princeton Math. Series 39, (1990).

[3] **D Gabai**, *The 4-dimensional Light Bulb Theorem*, Preprint arXiv:1705.09989v2 (2017).

[4] **F Laudenbach** (1973). *Sur Les 2-Spheres D'une Variete de Dimension 3*, Annals of Math. 97(1) (1973) 57–81.

[5] **R Norman**, *Dehn's lemma for certain* 4*–manifolds*, Invent. Math. 7 (1969) 143–147.

[6] **K Reidemeister**, *Zur 3-dimensionalen Topologie*, Abh. Math. Sem. Univ. Hamburg 11 (1933) 189–194.

[7] **R Schneiderman**, **P Teichner**, *The group of disjoint 2–spheres in 4–space*, Ann. of Math. Nov 2019 Vol. 190 no. 3, and arXiv:1708.00358v2 [math.GT] (2017).

[8] **R Schneiderman**, **P Teichner**, *Homotopy versus Isotopy: Surfaces in 5–manifolds*, Preprint in preparation.

[9] **H Schwartz**, *Equivalent non-isotopic spheres in* 4*–manifolds*, Preprint. arXiv:1806.07541v1 [math.GT] (2018).

[10] **J Singer**, *3-dim. manifolds and their Heegaard diagrams*, Trans. A.M.S. 35 (1933) 88–111.

[11] **R Stong**, *Uniqueness of $\pi_1$-negligible embeddings in* 4*–manifolds: A correction to theorem 10.5 of Freedman and Quinn*, Topology Vol 32 No 4 (1993) 677–699.

[12] **F Waldhausen**, *Heegaard-Zerlegungen der 3-Sphäre*, Topology 7, (1968), 195–203.

[13] **JHC Whitehead**, *A certain exact sequence*, Annals of Math. 52, (1950), 51–110.

[14] **H Whitney**, *The general type of singularity of a set of $2n-1$ smooth functions on $n$ variables*, Duke Journal of Math. 45, (1944), 220–293.



*E-mail address*: `robert.schneiderman@lehman.cuny.edu`

Dept. of Mathematics, Lehman College, City University of New York, Bronx, NY

*E-mail address*: `teichner@mac.com`

Max-Planck-Institut für Mathematik, Bonn, Germany